DETECTION OF CRUCIFORM DNA *IN VIVO*

THORSTEN ALLERS

Thesis presented for the Degree of Doctor of Philosophy
University of Edinburgh
1993

Abstract

Although cruciform extrusion by palindromic sequences in negatively supercoiled DNA is thermodynamically favoured, *in vitro* studies have suggested that it may be kinetically forbidden under physiological conditions. Only $(A\text{-}T)_n$ sequences that present no kinetic barrier to extrusion and palindromes flanked by (A+T)-rich DNA (which may extrude by an unusual pathway) have been shown to form cruciform structures *in vivo*. Nevertheless, long DNA palindromes of a normal base composition cannot be propagated in wild-type *Escherichia coli*. The palindromic sequence is either lethal to its vector or undergoes frequent deletion by replication slippage. The inviability of bacteriophage λ derivatives carrying long palindromes, which is associated with the inhibition of DNA replication, is overcome in *Escherichia coli sbcC* or *sbcD* mutant hosts. These mutants have therefore permitted the investigation of cruciform formation by long DNA palindromes *in vivo*.

This work describes an internally controlled, non-invasive assay for cruciform DNA *in vivo*, which relies on the inhibition of methylation at target sites that are in regions of unusual DNA secondary structure. If a palindrome with a central GATC target site for Dam methylase extrudes as a cruciform structure, then the GATC sequence will be located in a single-stranded loop and will consequently not be modified. The centre of a 476 bp perfect palindrome located in a bacteriophage λ derivative is shown to adopt a methylation-resistant DNA structure that is consistent with cruciform formation *in vivo*. Changes to the central base pairs are found to affect methylation inhibition, arguing that this extrusion occurs by a centre-dependent pathway. Using kinetic modelling, it is estimated that the palindrome centres examined are inaccessible to Dam methylase for between 35% and 88% of the λ lytic lifecycle.

A palindrome with an asymmetric insertion of 10 bp located in a bacteriophage λ derivative is found to exhibit unexpectedly high levels of methylation inhibition at a central GATC target site. In contrast, the λ phage carrying this inverted repeat shows a plating behaviour that is consistent with previous observations, which indicate that central asymmetry permits efficient replication by disfavouring cruciform extrusion *in vivo*. Furthermore, the viability (as measured by plaque areas on an *Escherichia coli sbcC* lawn) of the λ vectors carrying perfect palindromes is not commensurate with the degree of methylation inhibition at their palindrome centres *in vivo*. An attempt is made to reconcile these disparate results within the framework of current hypotheses for palindrome-mediated replicon inviability.

Acknowledgements

I would like to thank first and foremost my parents for their immeasurable help and support, and most recently for the provision of a computer which has facilitated all aspects of this work. I am also grateful to my brother Lars for quasi-spiritual advice.

I am indebted to my supervisor David Leach for scientific guidance; to Ewa Okely for technical advice and help throughout; to Elizabeth di Capua, Alison Chalker and Graham Bell for valuable support in the early days; and to John Connelly, Catherine Blake, and in particular to Angus Davison for easing the final passage of this work by their numerous words of wisdom and their enthusiastic proof-reading.

I would also like to thank Noreen Murray for the gift of λ and *E. coli* strains, and expert counsel in the ways of phage; Bob Lloyd for providing *E. coli* strains and their genealogy; Brenda Grimes and Howard Cooke for the use of a PhosphorImager; Chris Jeffree for generous help with image analysis and plaque size quantification; Gerry Smith for invaluable assistance in the search for $\chi^{+}A$; Waclaw Szybalski and Naomi Franklin for advice about N-mediated antitermination; and my advisor, Adrian Bird, for helpful discussions.

For much sympathy, advice, encouragement, and inebriation, I would also like to thank Patrick Costello, Roger Slee and Richard Smith.

Lastly, I would like to express my deep gratitude to Susan for a multitude of things, but above all for a sense of perspective.

Abbreviations

AAV	Adeno-associated virus
$Amp^{r/s}$	Ampicillin resistant/sensitive
AMPS	Ammonium persulphate
ATP	Adenosine-5'-triphosphate
att	λ attachment site
BIME	Bacterial interspersed mosaic element
bp	Base pair(s)
BSA	Bovine serum albumin
°C	Degrees centigrade
cAMP	3',5'-cyclic adenosine monophosphate
$Chl^{r/s}$	Chloramphenicol resistant/sensitive
CIP	Calf intestinal alkaline phosphatase
cm	Centimetre(s)
cos	Cohesive λ termini
χ	*chi* site
(d)dATP	2'(3'-di)-deoxyadenosine-5'-triphosphate
(d)dCTP	2'(3'-di)-deoxycytidine-5'-triphosphate
(d)dGTP	2'(3'-di)-deoxyguanosine-5'-triphosphate
(d)dTTP	2'(3'-di)-deoxythymidine-5'-triphosphate
(d)dNTP	2'(3'-di)-deoxynucleoside-5'-triphosphate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DTE	Dithioerythritol
DTT	Dithiothreitol
ΔG	Gibbs free energy change
$\Delta G^{\ddagger}$	Gibbs free energy of activation of transition state
ΔLk	Linking difference
EBNA	Epstein-Barr virus nuclear antigen
EBV	Epstein-Barr virus
EDTA	Diaminoethanetetra-acetic acid
ERIC	Enterobacterial repetitive intergenic consensus
FB	Foldback transposon
g	Gram(mes)
γ	*gam* protein of λ
HMG	High mobility group protein
HSV	Herpes simplex virus
IHF	Integration host factor
in	Inch(es)
IRU	Intergenic repeat unit
IS	Insertion sequence
J	Joule(s)
K	Kelvin
kb	Kilobase pairs

kcal ..... Kilocalorie
kDa ..... KiloDalton
Klenow ..... Large fragment of DNA polymerase I
krpm ..... Kilorevolutions per minute
l ..... Litre(s)
lb ..... Pound(s)
LE ..... Left end of λ
Lk ..... Linking number
λ ..... Bacteriophage lambda
m ..... Metre
M ..... Molar
mg ..... Milligram(s)
ml ..... Millilitre(s)
mm ..... Millimetre(s)
mM ..... Millimolar
MOI ..... Multiplicity of infection
mol ..... Mole(s)
MOPS ..... Morpholinopropanesulphonic acid
mRNA ..... Messenger ribonucleic acid
MVM ..... Minute virus of mice
μCi ..... Microcurie(s)
μg ..... Microgram(s)
μl ..... Microlitre(s)
μm ..... Micrometre(s)
μM ..... Micromolar
ng ..... Nanogram(s)
nm ..... Nanometre(s)
NMR ..... Nuclear magnetic resonance
OD ..... Optical density
orf ..... Open reading frame
ori ..... Origin of replication
^{32}P ..... β-emitting isotope of phosphorous
PEG ..... Polyethylene glycol
% ..... Percentage
PFU ..... Plaque forming units
pH ..... $-\log_{10}[H^+]$
pmol ..... Picomole(s)
Pol I ..... DNA polymerase I
Pol III ..... DNA polymerase III
PU ..... Palindromic unit
R ..... Gas constant (8.314 J K^{-1} mol^{-1})
RE ..... Right end of λ
REP ..... Repetitive extragenic palindrome
RF ..... Replicative form
RNA ..... Ribonucleic acid
RNase ..... Ribonuclease

s ..... Second(s)
^{35}S ..... β-emitting isotope of sulphur
SAM ..... S-adenosylmethionine
SDS ..... Sodium dodecyl sulphate
SSB ..... Single-stranded DNA binding protein
SV40 ..... Simian virus 40
σ ..... Superhelical density
$t_{½}$ ..... Half-time of reaction
TEMED ..... N-N-N'-N'-tetra-methyl-1,2-diamino-ethane
$Tet^{r/s}$ ..... Tetracycline resistant/sensitive
T_{MN} ..... Dinucleotide stability constant
Tn ..... Transposon
Tris ..... 2-amino-2-(hydroxymethyl)-1,3-propandiol
Triton X-100 ..... Octylphenoxypolyethoxyethanol
Tw ..... Helical twist
U ..... Unit(s)
UV ..... Ultraviolet
v ..... Version
V ..... Volt(s)
v/v ..... Volume per volume
v/w ..... Volume per weight
W ..... Watt(s)
Wr ..... Helix axis writhe
w/v ..... Weight per volume
w/w ..... Weight per weight
Xgal 5-bromo-4-chloro-3-indolyl-β-D-galactoside

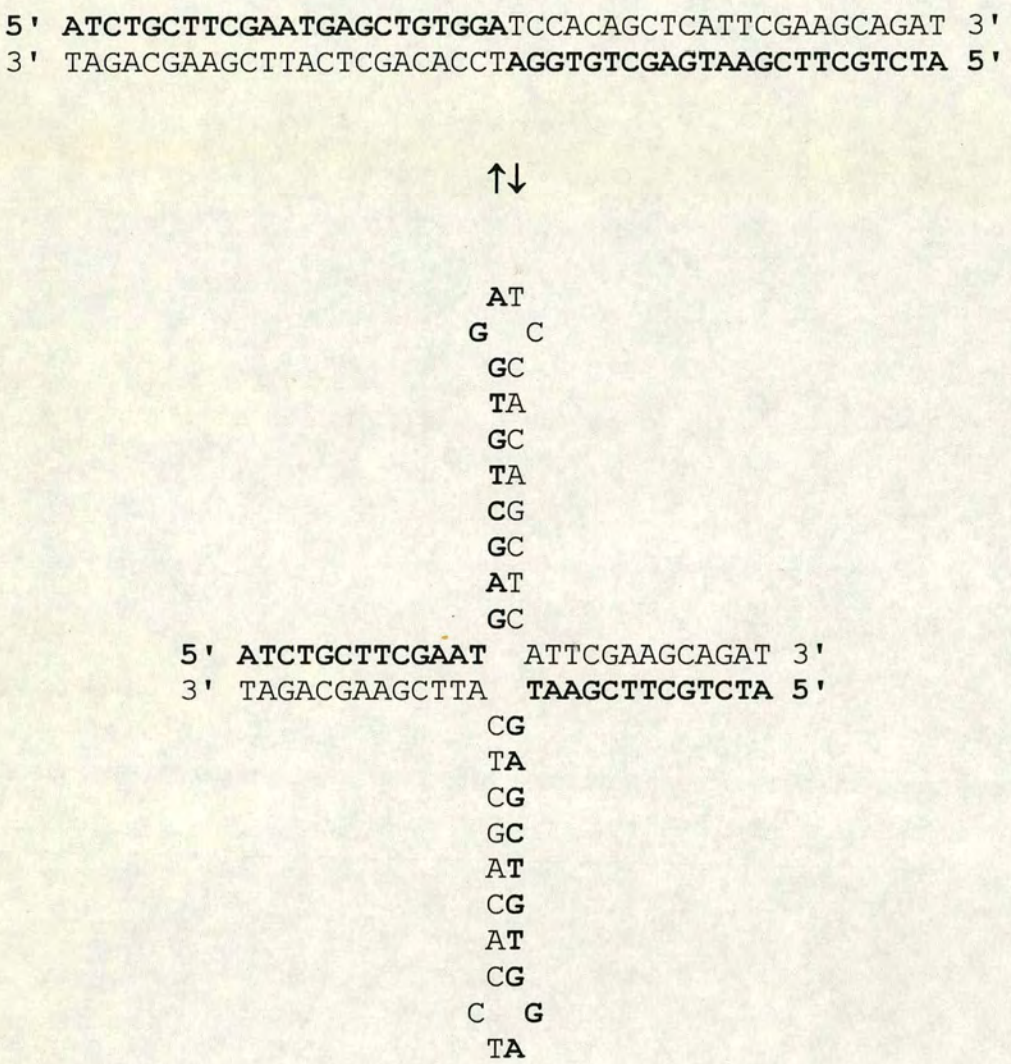
Table of Contents

CHAPTER 1

INTRODUCTION

Figure 1.1

Partial extrusion of a cruciform structure from a palindromic DNA sequence.



INTRODUCTION

A palindrome is a word or phrase that reads the same backwards as forwards, and has remained a poetic distraction for many hundreds of years. However, only two years after the proposal of the Watson and Crick model of DNA structure (Watson & Crick, 1953), Platt (1955) hypothesised that self-complementary DNA strands of an inverted repeat sequence could rearrange by a "transfer-twist" mechanism to form a four-way DNA junction. These inverted repeats have become known as palindromic DNA sequences, while the four-way junctions they may produce are called cruciform structures; the transfer-twist, or extrusion, is illustrated in Fig. 1.1.

In the following thirty-eight years palindromic DNA sequences of varying length and composition have been discovered in virtually all organisms examined for their presence. At the same time, cruciforms and similar structures have been implicated in numerous aspects of nucleic acid metabolism, such as transcription, translation, replication, chromatin assembly and most notably recombination (Holliday, 1964; Sigal & Alberts, 1972). However, the elusive and transitory nature of this unusual DNA secondary structure has resulted in cruciform DNA being largely refractory to detection and analysis *in vivo*.

This problem was compounded by the finding that in *Escherichia coli* long perfect palindromic sequences are either lethal to the vector or suffer frequent deletions that remove the symmetry (Collins, 1981; Lilley, 1981a). It was recently found that the mutation of a single gene, namely *sbcC*, could overcome this inviability (Chalker *et al.*, 1988); this property of *sbcC* mutants has enabled research into the behaviour of long palindromic sequences in *E. coli*. In particular, hypotheses invoking the extrusion of cruciform structures from palindromic DNA *in vivo* have received much attention, but their investigation was until lately hindered by the inability to detect such structural microheterogeneity in a non-invasive and reversible manner.

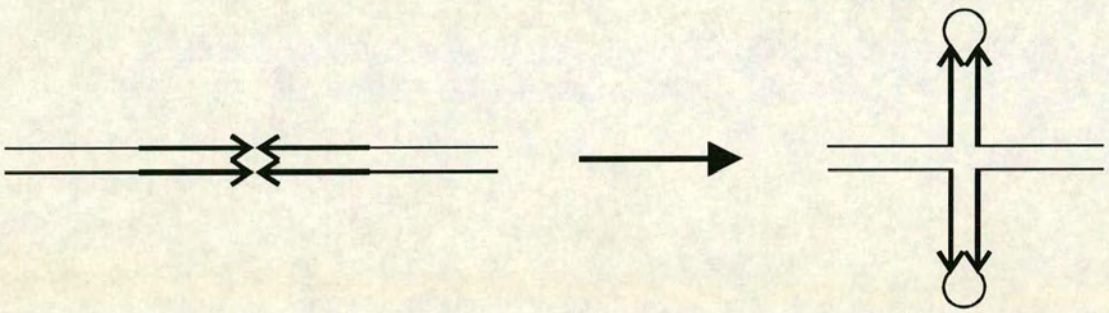
This thesis describes the application of a recently developed technique for characterising non-B-form DNA structures (Zacharias, 1992) to the vexed question of whether cruciforms exist *in vivo*. The assay used in this study employs the *E. coli* Dam methylase as an endogenous probe of the DNA secondary structure transition that is concomitant with cruciform extrusion. Research into the determinants of cruciform formation and the biological consequences of long DNA palindromes in *E. coli* was also undertaken. The results point unequivocally to cruciform extrusion *in vivo*.

SECTION 1

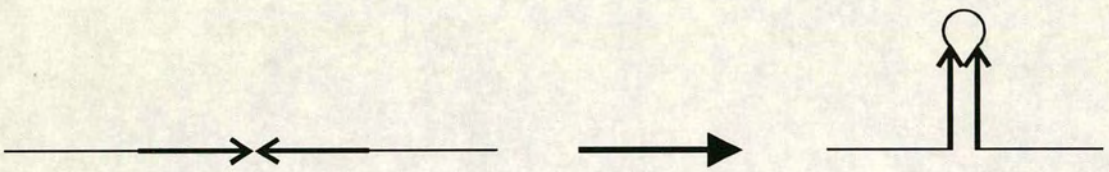
PALINDROMIC DNA

Figure 1.2

a. Extrusion of a cruciform from a double-stranded DNA palindrome.



b. Formation of a hairpin from a single-stranded DNA palindrome.



The term "DNA palindrome" was introduced by Wilson & Thomas (1974) to describe inverted repeats of the type

```
5' CGAATGAGCTGTGGATCCACAGCTCATTCG 3'
3' GCTTACTCGACACCTAGGTGTCGAGTAAGC 5'
```

where, unlike a literary palindrome, each DNA strand does not read the same backwards as forwards. This is due to the DNA palindrome having a two-fold rotational symmetry, rather than the mirror symmetry of its more poetic counterpart. However, given the antiparallel nature of double-stranded DNA, a polymerase reading such a palindrome would encounter the same sequence regardless of which strand it copied; to the polymerase, such a sequence would be palindromic. These inverted repeats are the subject of the first section of this introduction, while the cruciform structures they can form are reviewed in the second section. Although many of the topics discussed may overlap, they have been segregated depending on whether the study demonstrated, beyond reasonable doubt, the presence of a cruciform or hairpin. The distinction between cruciforms (which extrude from double-stranded DNA) and hairpins (which form by the self-association of single-stranded DNA) is illustrated in Fig. 1.2.

Naturally Occurring Palindromic DNA

Early Studies

Palindromic DNA sequences were discovered before the advent of DNA sequencing (Maxam & Gilbert, 1977; Sanger *et al.*, 1977) in a series of studies using electron microscopy and hydroxyapatite chromatography. The latter technique involves the denaturation of DNA with strong alkali, reneutralisation at high dilution to promote intramolecular associations, and adsorption to hydroxyapatite; subsequent elution with increasing concentrations of phosphate buffer will effect a chromatographic separation of single-stranded and double-stranded DNA (Wilson & Thomas, 1974). The first report of a rapidly self-associating single-chained structure, and therefore an unwitting discovery of palindromic DNA, was made by Bernardi (1962) using a similar technique to study chicken erythrocyte and calf thymus DNA.

Subsequent studies characterised this stable fraction more extensively (Walker & McLaren, 1965), but the work of Wilson & Thomas (1974) proved to be exemplary in its demonstration that the renatured DNA recovered by hydroxyapatite chromatography was composed of hairpin-like structures, that can originate from

inverted repeat, or palindromic sequences. The authors used techniques including electron microscopy and single-strand-specific nuclease digestion in conjunction with hydroxyapatite chromatography to examine DNA from *Drosophila melanogaster*, *Triturus viridescens*, *Xenopus laevis*, mouse and human cells. They concluded that the hairpins in their study did not have a detectable single-stranded loop, and were composed of perfect or near-perfect inverted repeats of a typical base composition, that were between 300 and 1200 nucleotides in length. While the distribution of the palindromic sequences was found to be species-specific, clustering of the inverted repeats was found in all the organisms studied.

Subsequent investigations extended this methodology: Schmid *et al.* (1975) found approximately 2000 – 4000 palindromes in the *Drosophila melanogaster* genome, which ranged in size from very short to greater than 15 kb (kilobase pairs). These inverted repeats differed from those found by Wilson & Thomas (1974) in that 80% of the palindromes in *Drosophila* appeared to have central spacers of between 0.5 kb and 30 kb, and no clustering of sequences could be found; a large proportion of middle repetitive DNA was also found in the 20% of hairpins with very small loops. The authors conclude their discussion with the caveat that contradictions between their work and that of Wilson & Thomas (1974) may have arisen out of the differing techniques for selecting inverted repeat sequences.

In a similar study Perlman *et al.* (1976) found that the genome of *Xenopus laevis* contains around 100,000 palindromes that appear to be inserted at different sites in different copies of the genome: renaturation kinetic studies showed that the unique DNA attached to the isolated palindromes could include all the single-copy sequences in the genome, and not just the small specific fraction that would be expected if the inverted repeats were always found at the same locations. The authors speculate that these palindromes may constitute another example of mobile DNA.

Renaturation kinetics, along with restriction endonuclease mapping, was also used by Engberg *et al.* (1976) and Karrer & Gall (1976) to show that in *Tetrahymena pyriformis* macronuclei, the free ribosomal DNA exists as perfect palindromes of over 20 kb in length; each palindromic molecule consists of two inverted copies of each of the two RNA genes. This arrangement appears to be unique to *Tetrahymena*, in common with many other aspects of its DNA organisation. Palindromic sequences have also been found in *Isospora (Toxoplasma) gondii*, a unicellular parasite (Borst *et al.*, 1984). Cruciform structures in circular mitochondrial DNA were observed by electron microscopy, indicating the presence of long inverted repeats.

While they may have identified palindromic DNA, the methods described above could not always provide unambiguous answers about its nature and occurrence. However, they clearly showed that while long palindromes occur frequently in the genomes of eukaryotes, they are totally absent in prokaryotes; this dichotomy in genomic organisation has prompted much speculation about the role of palindromes. As work proceeds to sequence the genomes of some of the organisms mentioned above, much of this speculation may be laid to rest, while newly discovered palindromes may pose fresh questions. In particular, inverted repeats that are too small or too imperfect to be detected by hydroxyapatite chromatography or electron microscopy have been found in the regulatory sequences of all organisms, where they may play an important part in DNA-protein interactions. This universal role for palindromes is discussed below.

Role of Palindromes in Transcription

Transcription Promoters

Protein-DNA interactions are dictated by the cognate binding of recognition motifs to their appropriate target sequences. Such DNA-binding domains are frequently symmetrical or exist as dimers, so it is hardly surprising that their target sites should exhibit a palindromic structure. One need look no further than the recognition sequences of commercially available restriction endonucleases to find a good example of dyad symmetry in protein-DNA interactions. The value of such an arrangement was reiterated very recently by Park *et al.* (1993) in a study involving concerted changes to the basic DNA-binding region of the v-Jun leucine zipper gene-regulatory protein and to its palindromic target site. The authors conclude that palindromic dimer binding sites provide a good design for selective molecular recognition: co-operation between the protein monomers in binding the target DNA increases the specificity of the protein-DNA interaction.

A statistical analysis of the significance of symmetrical sequences at protein-DNA interaction sites was performed by Dykes *et al.* (1975), who concluded that many of the known palindromic sites would be expected to occur by chance alone, and should therefore not be assumed to be involved in the interaction process. However, a search by Müller & Fitch (1982) for inverted repeats in the genomes of a wide variety of viruses showed a non-random distribution. Perfect or interrupted palindromes with repeat lengths of six or more base pairs occur more often than chance would predict, and are over-represented in regions that have a regulatory

significance. These palindromic sequences could represent binding-sites for dimeric proteins such as transcription factors, or as first suggested by Gierer (1966) could fold back on themselves to provide stem-loop structures. In this case the branched secondary structure, rather than the symmetrical primary structure, would be the target site for regulatory proteins. Subsequent analyses of operator sequences have shown that both modes of protein-DNA interaction are used in the regulation of gene expression, and selected examples are given below. Where appropriate, they are also discussed in the section on cruciform DNA.

The sequence analysis of regulatory regions from a wide variety of sources established that promoters are frequently associated with palindromes. In one of the first studies of this kind (Walz & Pirrotta, 1975) an inverted repeat was found at the P_R promoter of bacteriophage λ. The primary structure of this sequence, consisting of a mirror symmetry of 5 bp (base pairs) flanked by inverted repeats of 5 bp each, suggests a dimeric protein binding site. This turned out to be the case, for shortly afterwards it was identified as part of the 17 bp O_R2 binding site for λ *c*I repressor (Humayun *et al.*, 1977) and *cro* protein (Johnson *et al.*, 1978), both of which bind their targets as dimers.

A near-perfect palindrome of 18 bp total length found at the transcription start site of chicken mitochondrial DNA (L'Abbé *et al.*, 1991) is noteworthy in that it is located within a bidirectional promoter. If transcription were to proceed simultaneously in both directions then the palindrome would be located in a highly negatively supercoiled domain (Liu & Wang, 1987), and its extrusion as a cruciform would therefore be energetically favoured. The thermodynamics and consequences of such a rearrangement, and the twin supercoiling domain model of Liu & Wang (1987) will be discussed in the section on cruciform DNA.

A similar arrangement has been found at the *mob* promoter sequence of the *Thiobacillus ferrooxidans* plasmid pTF1 (Drolet & Lau, 1992). The 23 bp near-perfect palindrome located within the bidirectional *mob* promoter is also positioned in the overlapping minimal *oriT*, the origin of conjugative DNA transfer. Only when the *oriT* region is allowed to form a hairpin structure *in vitro* does it specifically bind MobL, one of the gene products transcribed from the *mob* promoter. It is therefore possible that the transcription-driven negative supercoiling of the *mob* promoter facilitates the extrusion of a cruciform at *oriT*, which leads to the binding of MobL and ultimately to the initiation of conjugative DNA transfer.

An inverted repeat sequence with 7 bp repeats was found in the promoter region of the *gyrA* gene of *Klebsiella pneumoniae* (Dimri & Das, 1990). The palindrome lies between the consensus –10 and –35 nucleotide sequences, and if extruded as a cruciform *in vivo* would almost certainly interfere with promoter recognition and the initiation of transcription. It is therefore particularly relevant that the potential cruciform resembles a rho-independent terminator, having a G/C-rich stem followed by an A/T-rich sequence. The stimulus for extrusion could be provided by the product of the *gyrA* gene, DNA gyrase, which introduces the negative superhelical turns necessary for cruciform formation. Homeostatic control of DNA gyrase would therefore be mediated by the palindromic sequence in its promoter.

Transcription Terminators

Palindromes are intimately associated with regions of transcription termination by virtue of their ability to form RNA hairpins when transcribed (see review by Platt, 1986). Although they constitute a well-documented example of hairpin structure, transcription terminators cannot be considered as cruciform DNA and will consequently not be discussed in those terms. On the other hand, the palindromic sequences in transcription terminators are described below.

Rho-independent termination at simple prokaryotic terminators depends primarily on two structural features: a region of G/C-rich inverted symmetry in the DNA preceding the transcription stop site and a series of uridines at the 3' end of the transcript. *In vitro* transcription analysis has established that the palindromic sequence is essential for termination: the formation of intramolecular hairpin structures in the RNA transcript elicits the polymerase pausing that is an obligatory prelude to rho-independent termination (Farnham & Platt, 1981). The palindromes in question generally consist of highly complementary inverted repeats with a repeat length of around 10 bp.

Attenuation of transcription, a special form of termination, is used in the regulation of gene expression and has been studied extensively in *Enterobacteriaceae*. Attenuation requires transcription by RNA polymerase of a leader region of 100 – 200 bp preceding the first initiation codon of a polycistronic gene cluster . In the case of the *Escherichia coli trp* operon (Bertrand *et al.*, 1975), the 162 bp leader contains many inverted repeat structures that closely resemble rho-independent transcription terminators. The RNA transcript of this attenuator may adopt a number of secondary structures (Lee & Yanofsky, 1977) that have the potential to terminate transcription, and thereby regulate gene expression. This is accomplished by four regions of inverted

symmetry that may form a number of mutually exclusive RNA hairpins (Oxender *et al.*, 1979). Intracellular levels of tryptophan indirectly determine the formation of RNA secondary structure, and consequently regulate transcription of the tryptophan operon (Oxender *et al.*, 1979). This model for attenuation has been extended to the tryptophan operon of *Serratia marcescens* (Stroynowski & Yanofsky, 1982) and subsequently to many other transcriptional units (Platt, 1986), all of which require inverted repeats in the leader region for modulation of gene expression.

Rho-dependent terminators are less well understood, primarily because the few examples that have been identified exhibit only limited homology. Nevertheless, five tandem rho-dependent terminators have been identified in the λ rightward operon (which initiates from the P_R promoter), all of which exhibit some dyad symmetry (Morgan *et al.*, 1983a). RNA polymerase pausing is observed at inverted repeats that allow the formation of hairpins with stems of 7 – 10 bp (Morgan *et al.*, 1983b); many of these sequences coincide with sites of rho-dependent termination.

By comparison with prokaryotes, transcription termination in eukaryotes is a complex process involving many sequential steps in the processing of the 3' RNA terminus (Birnstiel *et al.*, 1985). However, the components directing the 3' end formation of the histone gene transcripts are relatively well understood, and include features reminiscent of prokaryotic rho-independent terminators. These include a near-perfect DNA palindrome of 16 bp total length that is essential for the generation of mature 3' termini of sea urchin H2A mRNA (Birchmeier *et al.*, 1983). In common with prokaryotic terminators the inverted repeat has the role of encoding an RNA hairpin, and it is this stem-loop structure that is ultimately responsible of the cessation of transcription.

Role of Palindromes in DNA Replication

Origins of Replication

Palindromic DNA is a common feature of origins of replication where it may assist in the initiation of DNA synthesis, and has been extensively characterised in mammalian viruses. In particular, polyomaviruses such as simian virus 40 (SV40) (Fiers *et al.*, 1978), BK virus (Yang & Wu, 1979) and JC virus (Frisque, 1983) share many structural features of their replication origins. These include a perfect palindrome of 17 bp (which is nearly identical in all three viruses) and two imperfect palindromes (inverted repeats) of 15 – 23 bp and 24 – 27 bp total length, which also display considerable similitude (Frisque, 1983). The second of these inverted repeats

lies at the centre of the origin of replication and includes the second T-antigen binding site of SV40 (Tijan, 1978). The effect of stem-loop structures emanating from this inverted repeat on the binding of the SV40 T-antigen has been investigated (Tenen *et al.*, 1983), and will be discussed in the section on cruciform DNA.

Palindromes are also encountered at the DNA replication origins of human herpesviruses. The herpes simplex virus type 1 (HSV-1) ori_{L1} replication origin harbours a 144 bp perfect palindrome that displays striking homology to ori_S, the other origin of HSV-1 (Weller *et al.*, 1985). An almost identical 136 bp near-perfect palindrome was found in ori_{L2} origin of replication of herpes simplex virus type 2 (HSV-2) (Lockshon & Galloway, 1986). In both cases it is possible to discount dimeric protein-DNA interactions as the reason for the conservation of inverted symmetry, for the protein complex would have to be at least as long as the palindrome. On the other hand, the adoption of a cruciform conformation by the ori_L would explain the paired symmetrical base sequence differences between ori_{L1} and ori_{L2} that preserve the self-complementarity of the replication origin. This hypothesis is reinforced by finding that the central region of both palindromes is composed of an alternating AT sequence that would facilitate cruciform extrusion (Murchie & Lilley, 1987); the kinetics of cruciform extrusion will be discussed in detail in the second section of this introduction.

The latent origin of plasmid replication (*oriP*) of Epstein-Barr virus (EBV), a human herpesvirus, contains a 65 bp region of dyad symmetry and around 20 tandem copies of a 30 bp near-palindromic repeat (Baer *et al.*, 1984). Both of these sequences have a high affinity for Epstein-Barr virus nuclear antigen (EBNA-1) (Jones *et al.*, 1989), which binds at a 12 bp palindromic consensus site contained within the 30 bp repeats. Nevertheless, the participation of alternative secondary structures in EBV DNA replication should not be ruled out: single-stranded regions have been detected in *oriP* sequences that are present within supercoiled plasmids (Orlowski & Miller, 1991). The cleavage by T7 endonuclease I at sites located within the elements of *oriP* described above indicates the presence of branched DNA structures such as cruciforms *in vitro* (de Massy *et al.*, 1987).

Palindromes are not confined to the replication origins of eukaryotic viruses: bacteriophage λ initiates DNA synthesis from a site that displays a surprising complexity of symmetry elements and other characteristic features, which allow the formation of a highly ordered structure. A series of interspersed inverted repeats to the left of the origin can be folded into a cloverleaf structure, while a near-perfect

palindrome of 28 bp total length on the right of the origin can assume a cruciform conformation (Hobom *et al.*, 1979). The bacteriophage ϕ80 origin of replication, while different in terms of DNA sequence, contains potential cloverleaf and cruciform structures that are very similar to those found in λ.

Prokaryotic plasmids may contain regions of symmetry at the replication origins. For example, the broad host-range plasmids R1162 initiates DNA synthesis from two highly conserved positions flanking 40 bp inverted repeats separated by a 40 bp spacer (Lin & Meyer, 1987). Similarly, the leading strand replication origin of the staphylococcal plasmid pT181 consists of two adjacent inverted repeat elements, which are involved in origin recognition by the initiator (RepC) protein (Wang *et al.*, 1993). The binding of RepC protein has been shown to enhance cruciform formation at the pT181 replication origin *in vitro*, and frequency of extrusion *in vivo* correlates with the efficiency of RepC utilisation (Noirot *et al.*, 1990).

The nucleotide sequence of the *Escherichia coli* replication origin is characterised by a high degree of repetitiveness due to both inverted and direct repeats (Meijer *et al.*, 1979; Hirota *et al.*, 1979). The palindromic sequences are generally short and have central interruptions, indicating that they are most probably sites of protein-DNA interaction. However, both sets of authors found that *dnaG* primase recognition sites were accompanied by potential hairpin or cruciform structures; a parallel arrangement exists at the origin of replication of bacteriophage G4 and λ. Some of the inverted repeats found in *E. coli* are well conserved in *Enterobacteriaceae* replication origins, and have also been found in the marine bacterium *Vibrio harveyi* (Zyskind *et al.*, 1983).

Initiation at the human mitochondrial origin of light-strand DNA replication (O_L) requires the presence of 11 bp inverted repeats (Hixson *et al.*, 1986). This origin is well conserved amongst mammalian mitochondrial DNAs and functions only when the template strand is rendered single-stranded by displacement due to heavy-strand synthesis; this implies the formation of a stem-loop structure by the inverted repeat sequence.

Origins of replication isolated from mammalian genomic DNA have also been found to harbour palindromic sequences (Bell *et al.*, 1991). However, as these origin-rich sequences (ors) were purified using anti-cruciform DNA monoclonal antibodies, this work and related findings will be discussed in the section on cruciform DNA.

Viral Telomeres

A number of viral genomes are characterised by large terminal palindromes of several hundred base pairs in length that have been implicated in DNA replication. The self-complementarity of these sequences gives them the potential to form hairpins at the viral DNA ends, where they would be able to initiate DNA synthesis of the 5' terminus by a self-priming mechanism; this possibility was first suggested by Cavalier-Smith (1974). The discovery of covalently linked terminal hairpin loops and flip-flop sequence inversions at the ends of mature vaccinia virus DNA molecules (Baroudy *et al.*, 1982) has lent weight to this model of viral DNA replication.

One of the best researched examples of palindromic viral telomeres is the minute virus of mice (MVM), an autonomous mammalian parvovirus. Sequence analysis of the virion and replicative forms of MVM (Astell *et al.*, 1983; Astell *et al.*, 1985) revealed long imperfect palindromes, and has led to the proposal of a "modified rolling hairpin model" for viral DNA replication. This model differs from those held previously in that it accommodates the absence of sequence inversion at the 3' telomere of MVM. The recent discovery of the asymmetric resolution of the 3' parvovirus palindrome by the MVM NS1 polypeptide *in vitro* (Cotmore *et al.*, 1993) broadly endorses the model of Astell *et al.* (1985).

The use of palindromes in replicating the 5' ends of linear DNA molecules may not be restricted to viruses: a terminal repeat sequence of 26 bp length has been located at the ends of small linear DNA molecules found in the macronucleus of the ciliated protozoan *Oxytricha* (Wesley, 1975; Lawn, 1977). Additionally, two inverted copies of this repeat sequence surround single-strand DNA nicks at internal locations in the linear molecule. This sequence arrangement suggests the self-priming mechanism for DNA replication mentioned above. Recent work on the structure of *Tetrahymena* and *Oxytricha* telomeres has revealed the formation of four-stranded DNA by dimerisation between hairpin loops (Sundquist & Klug, 1989; Kang *et al.*, 1992), reiterating the similarity to viral DNA replication.

Palindromic DNA in Mobile Genetic Elements

The role of terminal inverted repeats is not restricted to DNA replication, for such sequences have been found at the ends of nearly all mobile genetic elements. The terminal repeats of different mobile DNAs vary both in length and degree of sequence conservation, but their ubiquitous nature emphasises the central importance of inverted repeats to transposition. While perfect palindromes are rarely found in mobile

DNA, they can easily be created *in vitro* using the long terminal inverted repeats of bacterial transposons; these artificial palindromic sequences have been used to investigate the phenomena of instability and replicon inviability. A short discussion of the structure of these transposable elements will therefore assist the review of the behaviour of long DNA palindromes in *E. coli*.

Bacterial Insertion Sequences and Transposons

Bacterial insertion sequences (IS) are widely disseminated mobile elements that vary in length between 0.8 and 2.5 kb (reviewed by Galas & Chandler, 1989). Sequence analysis of IS*1*, the best characterised IS element, revealed that the central 704 bp encoding the transposase protein is flanked by almost perfect 23 bp inverted repeats (Ohtsubo & Ohtsubo, 1978); to the outside of these lie 9 bp duplications of the IS target site. This sequence arrangement has been found in the other known IS elements, where the inverted repeats are 8 – 40 bp and the target site duplication is 3 – 13 bp. Transpositional recombination is generally dependent on the integrity of both ends of the IS, which must be correctly oriented on the same DNA molecule, and almost certainly provide the recognition sites for the element's *cis*-active transposase protein (the mechanistic aspects of transposition are reviewed by Grindley & Reed, 1985).

Transposons differ from IS elements in that besides the one or two genes involved in transpositional recombination, they encode genetic markers such as drug resistance. The Tn*3* family of simple (class II) transposons encode two *trans*-acting gene products, which act in a two-step replicative mechanism that requires the presence of 35 – 40 bp terminal inverted repeats (reviewed by Sherratt, 1989). However, it is the structure of the composite (class I) transposons that is of greater relevance to this discussion; these consist of two copies of an insertion sequence surrounding a set of resistance genes. The tetracycline-resistance transposon Tn*10*, which is 9.3 kb in length, consists of a 6.7 kb central region flanked by two 1329 bp IS*10* elements in opposite orientations (see review by Kleckner, 1989). These IS elements thereby constitute large inverted repeats, which possess their own 23 bp inverted repeat termini and transposase gene; only $IS10_R$ (IS*10*-Right) encodes a functional enzyme. Composite transposons such as Tn*10* (and the IS*10* element) exhibit a non-replicative mode of transposition.

Another composite transposon that has been studied in some detail is Tn*5*, which is a 5.8 kb mobile element composed of a central 2750 bp segment encoding resistance to kanamycin, bleomycin and streptomycin, and terminal inverted repeats of

the 1533 bp insertion sequence IS*50* (reviewed by Berg, 1989). Like Tn*10*, Tn*5* transposes by a non-replicative mechanism involving the *cis*-acting transposase encoded by $IS50_R$, and inserts are bracketed by 9 bp direct repeats of the target sequence.

While the inverted repeats of Tn*5* are transposable and thereby constitute IS elements in their own right (Berg *et al.*, 1981), they also promote the excision of Tn*5* in a manner that does not lead to its transposition to new sites. However, the requirements for these two processes are very different: transposition may occur when the repeats are in a direct, rather than inverted orientation and is dependent on the transposase function encoded by $IS50_R$. Conversely, excision is dependent on the inverted repeat conformation, but not on the action of Tn*5* transposase or the *E. coli* RecA protein (Egner & Berg, 1981). The removal of the central region of Tn*5* to generate a perfect palindrome greatly stimulates this excision (Collins, 1981; Collins *et al.*, 1982); such deletion-derivatives of Tn*5* (and Tn*10*) have been used in further studies on *recA*-independent (illegitimate) recombination, and are discussed later within the context of palindromic DNA instability.

Palindromes are involved in the acquisition of antibiotic resistance genes by Tn*21* and related class II transposons, which occurs by site-specific integration at either one of two distinct genetic loci (Martinez & de la Cruz, 1990). These recombination hot-spots are located at the 3' termini of a loosely conserved palindromic sequence of about 59 bp, which flanks the inserted genes in a number of naturally occurring transposons. The integrase is Tn*21*-encoded and may recognise the palindrome as a linear or cruciform structure.

Bacterial DNA Inversion Systems

Phase variation in bacteria and host-range variation in several bacteriophage is characterised by invertible DNA segments (reviewed by Glasgow *et al.*, 1989). The system may involve the inversion of a promoter relative to a stationary gene (as in *Escherichia coli* and *Salmonella typhimurium* phase variation), or the reorientation of structural genes relative to a fixed promoter (as in the host-range variation of bacteriophage Mu and P1). These closely related mechanisms require endogenous *trans*-acting factors for inversion, and the target sites of these recombinases flank the invertible DNA segment. Molecular characterisation of the Hin-related inversion systems (Hin, Gin, Cin and Pin) has revealed a consensus 26 bp recombination site (Plasterk & van de Putte, 1984) that harbours 12 bp inverted repeats; these are required for efficient inversion by the respective recombinase.

Drosophila melanogaster FB Transposons

Eukaryotic transposons are considerably more diverse than their bacterial counterparts, and it is therefore difficult to give a generalised account of the role of palindromes in the mobile genetic elements of higher organisms. However, an exception should be made for the FB (foldback) transposon family of *Drosophila melanogaster* because of its unusual structure (see review by Bingham & Zachar, 1989).

The termini of FB transposons consist of very large inverted repeats that vary in size from a few hundred base pairs to several kilobases (Truett *et al.*, 1981), and have a complex, internally repetitious structure that is best illustrated by the fully sequenced FB4 element (Potter, 1982). The 1095 bp terminal inverted repeats of the FB4 transposon consist of three regions of tandem or interspersed direct repeats. The length of these repeats varies from 10 – 31 bp, depending on the region, and copies of the shorter repeats are found in the longer sequences. Whereas the central portions of FB elements are heterogeneous and in some cases do not exist at all, the termini display strong sequence conservation. Their divergent lengths arise primarily out of variation in the number of 31 bp tandem repeats. Little is known about the mechanisms of FB transposition, but the terminal inverted repeats are presumably involved in interactions with putative transposases.

Repetitive Palindromic DNA

Eukaryotic Satellite DNA

The concurrent absence of highly repetitive satellite DNA and long DNA palindromes in prokaryotes suggests that inverted repeat sequences could be a component of the "junk" fraction of eukaryotic genomes. Studies on the renaturation kinetics of palindromes in eukaryotic DNA (Walker & McLaren, 1965; Schmid *et al.*, 1975) indicated that they constitute a significant part of the middle and highly repetitive DNA. Subsequent DNA sequence analysis has unveiled the palindromic component of satellite DNA in many species; recent examples include plants such as *Tropaeolum majus* (Nagl *et al.*, 1991) and 27 different species of waterfowl (Anatidae) (Madsen *et al.*, 1992). The frequent discovery of satellite DNA rich in inverted repeats, such as in the hermit crab *Pagurus pollicaris* (Fowler & Skinner, 1985), has tempted many researchers to speculate on the possible role of palindromes in propagating extragenic sequences. The association may be far from coincidental, for mammalian origins of replication isolated by anti-cruciform DNA affinity

purification, and therefore containing palindromes, hybridise to both monkey α-satellite and human Alu sequences (Bell *et al.*, 1991).

Bacterial REP Sequences

DNA sequence analysis of the genomes of *Escherichia coli* and *Salmonella typhimurium* has revealed the presence of a family of repetitive extragenic palindromic (REP) sequences, which have also been termed palindromic units (PU) (Higgins *et al.*, 1982). REP sequences have a consensus length of around 40 bp and consist of highly conserved (but imperfect) G/C-rich inverted repeats separated by a variable A/T-rich spacer (Gilson *et al.*, 1987).

REPs have been detected in other members of *Enterobacteriaceae* using the *E. coli* sequence as a probe, but display considerable species specificity (Gilson *et al.*, 1990a). Furthermore, they can be divided into two groups corresponding to the *E. coli* and *S. typhimurium* consensus sequences. REP sequences may occur singly or in clusters of two or more (Gilson *et al.*, 1984) in which they rigorously alternate in orientation. A recent analysis of clusters of up to six REPs revealed a number of conserved sequences (Gilson *et al.*, 1991); each cluster is composed exclusively of a mosaic combination of REPs and other sequence motifs, and has been termed bacterial interspersed mosaic element (BIME).

Present at a frequency of at least several hundred per chromosome (equivalent to between 0.5% and 1% of the genome), REPs are confined to extragenic regions and are mostly found at the 3' termini of genes. In spite of their expedient sequence and location, these 3' terminal REPs do not generally act as transcription terminators (Stern *et al.*, 1984), but have been implicated in the stabilisation of mRNA by protecting it from 3' → 5' exonuclease degradation (Stern *et al.*, 1988). However, this is unlikely to be their *raison d'être* as this function does not require the high degree of sequence homology observed in REPs (Gilson *et al.*, 1987).

REP sequences have been identified as binding targets for a number of bacterial proteins. If taken in concert, their activities suggest the involvement of REPs in chromosomal structure and organisation. The first such protein to be determined was DNA gyrase (Yang & Ames, 1988), the enzyme responsible for negative supercoiling. Although it has a large number of functions, most of these can be directly attributed to the level of DNA supercoiling, which is regulated by the opposing actions of topoisomerase I and gyrase (for a short review see Drlica, 1990). DNA gyrase is itself homeostatically regulated by supercoiling (Menzel & Gellert,

1983). Although no consensus target site has yet been identified (Lockshon & Morris, 1985), the specific interaction with REPs may be a strong indication that these sequences are the physiological sites of gyrase action on the chromosome.

This hypothesis has been corroborated by the discovery that binding of gyrase to REP sequences is stimulated by the histone-like protein HU (Yang & Ames, 1990). HU is a small, basic, abundant DNA-binding protein that is well conserved among bacterial species and shares considerable amino acid sequence homology with *E. coli* IHF (integration host factor). Native HU is heterodimer of two highly homologous components, HU-1 and HU-2, and belongs to the group of bacterial histone-like proteins (reviewed by Drlica & Rouviere-Yaniv, 1987) by virtue of its ability to wrap DNA: topological studies have shown that HU compacts DNA by coiling it in left-handed toroidal supercoils on a tetrameric protein complex (Broyles & Pettijohn, 1986). HU has been found to affect a number of site-specific protein-DNA interactions (Drlica & Rouviere-Yaniv, 1987).

Yang & Ames (1990) found that the stimulatory effect of HU on REP-gyrase binding extended to an increased efficiency of supercoiling and an inhibition of gyrase-mediated DNA cleavage. This implies that HU binding to the REP-gyrase complex regulates the supercoiling of chromosomal domains, and it is this modulation in the level of superhelical tension that in turn affects the site-specific interactions of proteins with DNA. Each one of the 100 or so *E. coli* chromosomal domains conceivably carries at least one gyrase-binding site (Snyder & Drlica, 1979), a frequency that is reasonably consistent with the estimated number of REP elements in the genome. The gyrase-HU-REP interaction may therefore be instrumental in the regulation of chromosomal topology and organisation. A stimulatory effect on REP-gyrase binding was not experienced for protein H1 (also called H-NS), a histone-like bacterial protein that has been implicated in the regulation of gene expression by DNA supercoiling (see review by Higgins *et al.*, 1990b).

The scope of the DNA-binding properties of protein HU was recently widened by the discovery that HU interacts specifically with kinked DNA (Pontiggia *et al.*, 1993). This capacity for the recognition of sharp angles was detected by using a gel retardation assay to identify and purify proteins forming complexes with a four-way DNA junction. While HU protein binds to four-way junctions and promotes their cleavage by T4 endonuclease VII, the archetypal junction resolving enzyme (Mizuuchi *et al.*, 1982a), it also inhibits the formation of cruciforms from supercoiled plasmids. HU binding could inhibit cruciform extrusion by relieving negative supercoiling,

thereby decreasing the amount of free energy available for the transition. The properties described above are shared by the non-histone high mobility group protein 1 (HMG1), an abundant eukaryotic nuclear protein of uncertain function (Bianchi *et al.*, 1989). As these analogous proteins interact specifically with cruciform DNA they will be discussed in greater detail in section two. The consequences of this finding for the gyrase-HU-REP model outlined above could involve the binding of HU to REP cruciforms as a signal for gyrase action, or the localised constraint of negative supercoiling by HU protein to prevent extrusion of the REP gyrase target site.

Another REP-binding activity that has been established recently is that of DNA polymerase I (Pol I), which forms stable complexes with the palindromic sequence in a gel retardation assay (Gilson *et al.*, 1990b). Although the significance of this interaction has yet to be resolved, REP sequences could serve as preferred entry sites for Pol I, playing a role in secondary replication initiation, in specific polymerase pausing or in replication fidelity. Seven REPs, three consisting of a cluster of two REP sequences, have been found in a 136 kb segment including the *E. coli* origin of replication (Burland *et al.*, 1993). While this is an average frequency for REPs, their presence may nevertheless be required for efficient DNA replication.

Other Repetitive Bacterial DNA

REP sequences are not the only form of repetitive DNA that has been identified in bacteria. A 126 bp sequence that is restricted to the intergenic regions of polycistronic operons or the untranslated regions upstream or downstream of open reading frames has been discovered in *E. coli*, *S. typhimurium* and other enterobacterial species by Sharples & Lloyd (1990) and Hulton *et al.*, (1991). This sequence has been termed intergenic repeat unit (IRU) and enterobacterial repetitive intergenic consensus (ERIC) respectively. Its frequency in the genome, commensurate with its length, is less than that of REPs, and although the two groups of repetitive sequences share no similarity on a nucleotide level, both possess an inverted repeat structure. IRU/ERIC sequences represent the longest (albeit rather imperfect) palindromes identified in bacteria to date. They show a high degree of sequence conservation, and in certain cases where they differ from the consensus a compensating base change is seen in the complementary arm of the hypothetical stem-loop structure. The chromosomal location of IRU/ERIC sequences is species-specific but unlike REPs they do not occur in clusters. Although no clear function has yet been attributed to IRU/ERICs, it has been suggested that they represent a form of "selfish DNA" (Hulton *et al.*, 1991).

Behaviour of Long DNA Palindromes in *Escherichia coli*

Naturally occurring palindromes in *E. coli* are short and generally imperfect; none of the examples given above display uninterrupted symmetry for greater than ~40 bp. Long DNA palindromes therefore represent a foreign substrate for the enzymatic machinery of *E. coli*. This predicament results in the palindromic DNA being lethal to its vector or suffering deletions that destroy its symmetry (Collins, 1981; Lilley, 1981a; Collins *et al.*, 1982; Mizuuchi *et al.*, 1982b; Hagan & Warren, 1982; Hagan & Warren, 1983; and others); these phenomena have been termed inviability and instability respectively.

The difficulty in propagating long DNA palindromes is of particular relevance to the cloning of human and other eukaryotic sequences. The absence of palindromic structures amongst a recombinant series of plasmid fragment polymers was observed during the first use of cosmid cloning vehicles (Collins & Hohn, 1978). The ensuing prediction that the cloning of eukaryotic sequences containing long palindromes would be problematic has been borne out by numerous studies (e.g. Nader *et al.*, 1985; Wyman *et al.*, 1985; Donlon *et al.*, 1986; Wertman *et al.*, 1986; Wyman *et al.*, 1986); a significant proportion of the recombinant phage λ clones from eukaryotic genomic libraries fail to grow in a wild-type *E. coli* host, and these clones frequently contain palindromic DNA.

The analysis of these irksome sequences has been possible by the use of hosts mutant for the *recB*, *recC*, *sbcB* and *sbcC* genes (Collins *et al.*, 1982; Leach & Stahl, 1983; Chalker *et al.*, 1988). In particular, it is the mutation of the *sbcC* gene that primarily overcomes the inviability associated with palindromic DNA. The stable propagation of long palindromes in *E. coli* has facilitated research into the determinants of instability and inviability, and the results of these investigations are reviewed below. While these phenomena are sufficiently distinct to merit separate discussion, their intertwined relationship may make this segregation seem arbitrary; it is probable that they are linked by a common cause — the potential of palindromic sequences to form unusual DNA secondary structures.

Palindrome-Mediated DNA Instability

Extent of Palindrome-Mediated DNA Instability

The difficulties encountered when cloning novel sequences in *E. coli* are accepted by many researchers as no more than an occupational hazard. However, the concurrence of deletions and palindromic sequences has not escaped the attention of a number of vigilant groups. As mentioned above, Collins & Hohn (1978) reported the absence of palindromic structures in the first use of cosmid vectors; they also noted a high number of deletions among their isolates (four out of 52). Similarly, Casadaban & Cohen (1980) failed to recover inverted dimer clones of plasmid DNA fragments when analysing regulatory DNA sequences by gene fusion. Sadler *et al.* (1978) found that ligation of synthetic lactose operator sequences into plasmid vectors frequently resulted in tandem multiple insertions; transformation of *E. coli* by these constructs regularly led to the excision of operators which were arranged as inverted repeats. This was confirmed by Betz & Sadler (1981) using palindromic dimers of this lactose operator. The deletion of palindromes at a high frequency has also been observed in *Streptococcus* (Behnke *et al.*, 1979), indicating that this phenomenon is not restricted to *E. coli*.

The first detailed analysis of the instability of palindromic DNA in *E. coli* was made by Collins (1981) using a deletion derivative of Tn*5*. The introduction of Tn*5* into the β-lactamase gene of pBR322 (Bolivar *et al.*, 1977b) and the subsequent removal of the central region of the transposon by restriction endonuclease cleavage generates a long palindromic sequence that undergoes frequent deletion in *E. coli*; later studies into the determinants of palindrome-mediated DNA instability have often used similar constructs based on the large inverted repeats of composite (class I) transposons.

The instability (and inviability) of a palindromic sequence of 260 bp was noted during early research into the formation of cruciform structures *in vivo* (Lilley, 1981a). Other groups working in this field have reported similar observations: Mizuuchi *et al.* (1982b) attempted the transformation of *E. coli* with a palindromic pBR322 fragment dimer that had been used in kinetic studies on DNA supercoiling and cruciform extrusion, but found that it was inefficient and resulted in deletions destroying the inverted symmetry. These rearrangements were not found when the pBR322 copies were separated by 300 bp of non-palindromic DNA. The instability of a 68 bp perfect palindrome in a pBR322 derivative was observed by Courey & Wang (1983) in a related investigation.

Large perfect palindromes also undergo frequent deletions in both double-stranded DNA phage such as λ (Perricaudet *et al.*, 1977; Leach & Stahl, 1983) and single-stranded DNA phage such as ϕX174 (Müller & Turnage, 1986). This is of particular concern to researchers wishing to construct recombinant DNA libraries from eukaryotic organisms in bacteriophage vectors. A genomic library of the plasmodial slime mould *Physarum polycephalum* cannot be efficiently propagated in a wild-type *E. coli* host. In particular, the actin-related clones are characterised by sequence instability, and electron microscopy of DNA fragments from these phage has revealed stem-loop structures that suggest the presence of long DNA palindromes (Nader *et al.*, 1985). A recombinant library prepared from the inverted duplicated human chromosome 15 isolated from Prader-Willi syndrome patients also includes unstable clones that contain long palindromes (Donlon *et al.*, 1986). The determinants of this sequence instability in *E. coli* may likewise cause the chromosomal aberrations associated with Prader-Willi syndrome; the possible involvement of unusual DNA secondary structures in genetic disease is discussed in the section on cruciform DNA.

Attempts to clone eukaryotic sequences in plasmid vectors are also plagued by the instability of long DNA palindromes. The 5' terminal palindrome of the minute virus of mice, which has been discussed previously in the context of DNA replication, cannot be stably propagated in wild-type *E. coli* and provokes site-specific deletions at the centre of symmetry (Boissy & Astell, 1985). Another sequence component of eukaryotic virus replication that undergoes frequent deletion in *E. coli* is the ori_L herpes simplex virus type 1 origin of DNA synthesis (Weller *et al.*, 1985); frustrated by the instability of ori_L, the authors resorted to the use of a yeast cloning vector. However, this approach may not be panacea sought by researchers keen to sequence the genomes of higher eukaryotes: a plasmid-borne palindrome of 94 bp is unstable in *Saccharomyces cerevisiae* (Henderson & Petes, 1993).

Determinants of Palindrome-Mediated DNA Instability

DNA palindromes do not delete of their own accord; while the sequence's length and composition may influence its own deletion, it is by no means the only determinant of instability. These factors are outlined below, followed by a review of the various hypotheses that purport to explain this phenomenon.

Palindrome Length

Early studies on the instability of palindromic DNA often employed very large inverted repeat sequences: Collins (1981) observed the deletion of a 976 bp Tn*5*

derivative, and Mizuuchi *et al.* (1982b) found that a head-to-head plasmid dimer of several thousand base pairs is unstable in *E. coli*. Subsequent studies on cruciform extrusion that employed palindromes of 76 bp (Sinden *et al.*, 1983) and 114 bp (Gellert *et al.*, 1983) demonstrated that constructs of this reduced size are not prone to deletion. This is in contrast to the inviability and instability of a palindromic sequence of 260 bp (Lilley, 1981a).

A 147 bp perfect palindrome derived from DNA fragments containing the SV40 replication origin was found by Bergsma *et al.* (1982) to be stable in *E. coli*. Sequence analysis of the centre of symmetry revealed a high G/C content, which could impede the extrusion of a cruciform (the kinetics of this structural transition are discussed in section two). This led to speculation as to whether the stable properties of the SV40-derived construct could be conferred upon other palindromes. However, attempts to extend the inverted sequence homology by a total of 3034 bp were unsuccessful and resulted in extensive deletions of the host vector (Hagan & Warren, 1983).

Systematic research into the palindrome size necessary for deletion has established that sequences with a total length of 190 bp or less are stable, while those of 368 bp or greater suffer from instability (Yoshimura *et al.*, 1986). While such results point to a demarcation between stable and unstable palindromes that is based on size (with a cutoff at around 200 bp), it should be remembered that sequences of 146 bp and 147 bp that are stable in *E. coli* display intermediate properties such as a reduction in plasmid copy number due to multimerisation (Warren & Green, 1985); this effect was not seen with a palindrome of 114 bp. Furthermore, other factors such as the host genotype can significantly alter this size limit. This effect is discussed later with respect to replicon inviability.

While the stability of palindromic sequences in bacteriophage ϕX174 also shows size-dependence, this single-stranded vector appears less tolerant of inverted repeats than the plasmids described above: only sequences of 28 bases or less are completely stable, while inserts of 38 and 48 bases are deleted with a frequency commensurate to their length (Müller & Turnage, 1986). However, a more detailed study using palindromes of a lower G/C content (60% G/C as opposed to 80% G/C) determined that sequences of up to 48 bases are maintained stably in the ϕX174 genome (Williams & Müller, 1987). The authors propose that the genetic stability of these palindromes is inversely related to the thermodynamic stability of the hairpins they can form.

The deletion frequencies of double-stranded DNA palindromes ranging in size from 42 bp to 106 bp were analysed by Sinden *et al.* (1991), who found that both the length and base composition of the inverted repeats contribute to their instability. In particular, within the first 20 bp of a potential hairpin stem a higher rate of deletion correlates with an increased thermodynamic stability of base pairing.

A correlation of instability with palindrome length is also seen at smaller inverted repeat sizes (although the excision frequencies are orders of magnitude lower than those described above): elongation of a 22 bp sequence to 90 bp exacerbates deletion several-fold, but this is strongly affected by the sequence context of the insertion site (DasGupta *et al.*, 1987). The role of sequence context in determining palindrome instability is discussed below.

The influence of the palindrome's length on the endpoint of its deletion has been investigated by Weston-Hafer & Berg (1989). The enhanced instability of a 90 bp perfect palindrome over a 22 bp sequence is associated with the use of 4 bp direct repeats as deletion endpoints in preference to 9 bp or 10 bp repeats. A closer analysis of this phenomenon has shown that deletion frequency and deletion endpoint location depend differently on palindrome length: a significant increase in the former is only experienced by sequences of 32 bp or more, whereas no more than 18 bp of inverted symmetry are required to bias the choice of direct repeat endpoint (Weston-Hafer & Berg, 1991a).

Direct Repeats

As mentioned above, direct repeats act as deletion endpoints for palindrome-mediated instability. This was established by studies of the excision of composite (class I) transposons such as Tn*5*, which is functionally identical to palindrome deletion. Excision is dependent on the inverted repeats but not on the transposase function of Tn*5* or the host RecA protein (Egner & Berg, 1981), is stimulated by the removal of the central asymmetric region (Collins, 1981) and primarily uses direct repeats such as the 9 bp target site duplications as deletion endpoints. Three classes of deletion have been established for the Tn*10* transposon: precise excision of Tn*10*, nearly-precise excision of Tn*10*, and precise excision of the nearly-precise excision remnant (Foster *et al.*, 1981). All three types of event occur between 4 – 10 bp direct repeat sequences at or near the ends of the element in question, and are markedly reduced in frequency in *E. coli recB sbcB* cells (Collins *et al.*, 1982); the role of host mutations is discussed later.

The relative contributions of inverted and direct repeats to sequence instability have been estimated in *E. coli* (Brunier *et al.*, 1989; Weston-Hafer & Berg, 1989; Weston-Hafer & Berg, 1991a; Pierce *et al.*, 1991) and in *Bacillus subtilis* (Peeters *et al.*, 1988). The findings are broadly similar: the frequency of deletion between direct repeats is exponentially related to their lengths (Brunier *et al.*, 1989; Peeters *et al.*, 1988; Pierce *et al.*, 1991), but may be increased yet further by the presence of an intervening palindrome. However, this palindrome-mediated stimulation is specific to short direct repeats (Brunier *et al.*, 1989; Pierce *et al.*, 1991), and can promote the use of 4 – 5 bp sequences as deletion endpoints in preference to longer and normally more efficient 9 bp direct repeats (Weston-Hafer & Berg, 1991a).

Direct repeats may cause deletions without the assistance of a palindrome. Examinations of spontaneous mutations in the *E. coli lacI* gene (Farabaugh *et al.*, 1978; Schaaper *et al.*, 1986) and in a *lac I-Z* gene fusion (Albertini *et al.*, 1982) have identified directly repeated sequences as mutational hotspots in their own right. Nevertheless, some of the deletions analysed by Schaaper *et al.* (1986) were also associated with inverted repeats. Models for the participation of palindromic DNA sequences in such deletion mutagenesis generally involve the juxtaposition of direct repeat hotspots by a hairpin structure (Ripley, 1982; Ripley & Glickman, 1983; Glickman & Ripley, 1984); the importance of primary and secondary DNA structure in deletion between direct repeats has been reaffirmed by Trinh & Sinden (1993).

Deletions mediated by direct repeats alone can be just as extensive and frequent as those brought about by palindromes: the loss of 2.5 kb of plasmid DNA between 7 bp direct repeats has been reported by Jones *et al.* (1982), while the spontaneous deletion of 2.2 kb of M13 replicative form (RF) DNA harbouring fragments of the T4 *den*V gene is observed in 95% of the RF DNA within an infected cell (Lloyd & Augustine, 1987). Like the precise excision of transposons, this type of deletion is independent of the host *recA* gene, although in contrast to palindrome-mediated instability it is enhanced by *recB recC* mutations (Mazin *et al.*, 1991). The implications of such findings will discussed later with respect to the host-encoded determinants of sequence instability and replicon inviability.

Asymmetry at the Palindrome Centre

The presence of an asymmetric interruption at the centre of a palindrome can significantly restore its stability: removal of the 2750 bp central region of Tn*5* to generate a perfect palindrome stimulates the precise excision of this transposon by four to five orders of magnitude (Collins *et al.*, 1982). Conversely, when 42 kb of

phage λ DNA is inserted between two IS*50* elements, the Tn*5*-related construct is 50 – 1000 times more stable than a perfect palindrome that can be formed by removal of all but 32 bp of the insertion element terminal inverted repeats (DasGupta *et al.*, 1987).

The size of asymmetry need not be this large to stabilise a palindrome: a step-wise reduction (to 173 bp) in the length of a 1.4 kb spacer between inverted repeats derived from Tn*5* does not increase the frequency of deletion in *Bacillus subtilis* (Peeters *et al.*, 1988). A detailed study of the minimum asymmetry required to permit stable propagation was undertaken by Warren & Green (1985). While long palindromes with inserts of 57 bp and 72 bp can be recovered from a wild-type *E. coli* host, they remain unstable; only inserts of greater than 150 bp completely eliminate the sequence instability.

Sequence Context

The immediate environment of a DNA palindrome may influence its stability considerably. Insertion of palindromic sequences between 22 bp and 90 bp in length at 13 different sites in pBR322 markedly alters their deletion frequency: a 1 bp shift in position can stimulate the instability of the 90 bp inverted repeat 3000-fold (DasGupta *et al.*, 1987). This variation in deletion is not systematic with respect to location, and depends in part on the length of the palindrome. Similarly, different plasmids harbouring identical palindromes may experience deletion frequencies varying by 130-fold (Hagan & Warren, 1983).

Relocation of a palindromic sequence to different chromosomal contexts can also affect its stability. Kazic & Berg (1990) interrupted the β-lactamase gene (*bla*) of pBR322 with palindromes of 22 bp and 90 bp; deletion could then be detected by a reversion to Amp^r. The mutant *bla* genes were then placed in the same position in *lacZ*, and the constructs were moved to three different locations: to an F'*lac*, to a λ*lac* integrated in the *E. coli* chromosome at *att*λ, and to the *E. coli* chromosomal *lac* gene. Deletion of the palindromes was found to vary substantially, although for most sequences placement on larger, single-copy chromosomes reduced the frequency of reversion relative to pBR322, often by one or two orders of magnitude. Again, the magnitude of this effect is specific to both the insertion and location.

DNA Replication

The instability of palindromic DNA is intimately associated with its replication. When a λ phage with an 8.4 kb palindrome infects an *E. coli* rec^+ *dnaBts* strain and

DNA replication is prevented by growth at 42°C, there is no loss of the palindromic sequence or its carrier replicon. Conversely, when DNA synthesis is permitted at 37°C those phage that have undergone replication are characterised by extensive deletions of the inverted repeat sequence (Shurvinton *et al.*, 1987). The replication of a palindrome is therefore a prerequisite for its demise.

Palindromic sequences are renowned for their ability to arrest DNA synthesis (Weaver & DePamphilis, 1984). The authors found that sites in circular single-stranded DNA templates which stop replication by mammalian DNA polymerase α *in vitro* often consist of palindromes that form stable hairpin structures in solution; synthesis is arrested precisely at the base of the stem-loop structure. However, when the locations of arrest sites utilised *in vivo* by SV40 DNA replication were mapped, the participation of inverted repeats was not observed. This may not be unsurprising given the tolerance of palindromic DNA exhibited by eukaryotic organisms.

A similar analysis of the arrest of *in vitro* replication using *E. coli* DNA polymerase III (Pol III) was performed by LaDuca *et al.* (1983). Approximately 65% of the sites at which pausing was observed were within 15 nucleotides of potential template hairpins; the efficiency of the pause site and its proximity to the base of a stem-loop structure is affected by the subunit composition of the enzyme. The inclusion of *E. coli* single-stranded DNA binding protein (SSB) in the reaction removes virtually all barriers to the progression of the holoenzyme. However, the supply of SSB *in vivo*, which is limited to around 1000 tetramers (Meyer & Laine, 1990), may be insufficient to ensure the accurate replication of a long palindrome.

The ability of polymerase pausing to elicit deletion or frameshift mutations in the newly synthesised DNA strand has been examined *in vitro* by Papanicolaou & Ripley (1991). It was found that the termini produced by pausing correlate with repeated and palindromic or quasipalindromic sequences, which are likely to be precursors to mutagenic misalignments. The deletion of an inverted repeat during DNA synthesis is therefore dependent not only on its replication, but also on the pausing of the polymerase at that sequence.

Whether palindrome-mediated deletions occur preferentially on the lagging strand of replication is unclear. The Okazaki fragment-sized single-stranded regions of the lagging strand template would predispose it to secondary structure formation and misalignment, and consequently deletions of palindromic sequences would be more frequent on this strand. However, experiments to test this hypothesis have disagreed in their conclusions: Weston-Hafer & Berg (1991b) found that changing the

replication direction of palindromes between 22 bp and 90 bp in size did not alter the frequency of deletion or distribution of endpoints, while Trinh & Sinden (1991) found moving a 17 – 18 base asymmetric palindromic insert from the leading to the lagging strand stimulated mutagenesis 10 – 20-fold. These results may not be as contradictory as they seem: although both groups used reversion to drug resistance to assay deletion, Weston-Hafer & Berg (1991b) inserted their palindromic inserts into the *amp* gene of pBR322 (resulting in sensitivity to ampicillin), and Trinh & Sinden (1991) interrupted the *CAT* gene of pBR325 with their palindromic sequences (to give sensitivity to chloramphenicol). The use of chloramphenicol as a selective agent may have elicited aberrant DNA replication; this is outlined below.

In wild-type *E. coli*, replication of ColE1-related plasmids such as pBR325 is initiated by DNA polymerase I (Pol I), which extends an RNA primer, and is then continued by DNA polymerase III (Pol III) (Staudenbauer, 1976). When Pol III is inactivated *in vivo* by the use of a temperature-sensitive mutation, a 58 bp palindrome in pBR345 (consisting of two 29 bp *lac* operators) interferes with plasmid DNA synthesis by Pol I: replication intermediates containing hairpins accumulate at the centre of the sequence as a result of frequent strand switching (Backman *et al.*, 1978). Template switching of this kind is a pivotal component of hypotheses for palindrome-mediated sequence instability, which is elucidated later. Treatment of wild-type cells with chloramphenicol leads to equivalent levels of strand switching, and it has been suggested that the differential stability of Pol I and Pol III in the absence of protein synthesis lies behind this phenomenon (Bolivar *et al.*, 1977a). The preferential lagging strand mutagenesis observed by Trinh & Sinden (1991) may therefore be due to error-prone DNA replication by Pol I, that has been brought about by the use of chloramphenicol. While evidence has recently emerged suggesting that other forms of mutagenesis may also occur preferentially on the lagging strand (Veaute & Fuchs, 1993), an involvement of Pol I has not been excluded by the authors.

The polymerase-specificity of palindrome-mediated instability has been identified in *Saccharomyces cerevisiae*: mutations in *POL1*, the gene encoding the large subunit of yeast DNA polymerase I, increase the deletion frequency of 60 bp and 160 bp palindromic sequences (Ruskin & Fink, 1993). The excision in yeast of the bacterial transposon Tn*5* between its terminal 9 bp direct repeats is also stimulated by mutations in *POL1* or *POL3* (Gordenin *et al.*, 1992). Evidence has recently emerged that long inverted repeats can act as hotspots for mitotic interchromosomal recombination, possibly as a result of altered DNA replication at the base of a hairpin formed by the palindromic sequence (Gordenin *et al.*, 1993).

The generation of single-stranded DNA during replication stimulates the instability of palindromic sequences. Excision of Tn*5* is strongly enhanced when the transposon is inserted into F' episomes or Hfr chromosomes (Egner & Berg, 1981), and it has been shown that the increased deletion frequency is a consequence of DNA transfer in conjugation (Berg *et al.*, 1983). This F' plasmid stimulation of excision is negated by Tra^- mutations which are defective in conjugal transfer, and occurs in the recipient cell during mating (Syvanen *et al.*, 1986). It was proposed that replication of the single-strand of donor DNA in the recipient to generate a duplex molecule is responsible for the deletion. Evidence to support this hypothesis was subsequently provided by Brunier *et al.* (1988): induction of single-stranded DNA synthesis of a double-stranded pBR322-based plasmid stimulates the excision of a Tn*10*-derived transposon up to 10^6 times. Deletion between short direct repeats is also greatly enhanced by this mode of replication, but in keeping with findings outlined above, is increased yet further by the presence of inverted repeats (Brunier *et al.*, 1989). Palindromes are consequently expected to be unstable in single-stranded replicons. This has been shown for bacteriophage ϕX174, where palindromic sequences of more 28 bp are unstable (Müller & Turnage, 1986), and bacteriophage M13, into which 110 bp inverted repeat could not be stably cloned (Leach *et al.*, 1987).

Host Genotype

The instability of palindromic sequences in *E. coli* is independent of the *recA* gene (Collins, 1981), and can therefore be considered as illegitimate recombination (Franklin, 1967). However, the mutation of other genes involved in general (RecA-dependent) recombination and DNA repair can affect the frequency of palindrome-mediated deletions. Most notably, the inactivation of exonuclease V (by mutation of *recB* and *recC*) and exonuclease I (by mutation of *sbcB*) appears to allow the stable propagation of inverted repeats. The higher stability of palindromes in a *recB sbcB* host was first reported by Collins *et al.* (1982), and confirmed by Leach & Stahl (1983) using *recBC sbcB* cells. A genomic library of *Physarum polycephalum* that is rich in inverted repeats can only be propagated intact on a *recBC sbcB* host (Nader *et al.*, 1985). Boissy & Astell (1985) attempted to grow plasmids containing the 5'-terminal palindrome of minute virus of mice in such a strain, but found an additional *recF* mutation to be necessary for stability. Similarly, Yoshimura *et al.* (1986) used *E. coli recBC sbcB recF* to stably maintain long artificial palindromes. However, Lockshon & Galloway (1986) found a plasmid containing ori_{L2}, the large palindromic DNA replication origin of herpes simplex virus type 2, to be more stable in a *recBC sbcB* strain than a *recBC sbcB recF* strain.

While the *recF* mutation does not have a direct effect on the frequency of palindrome deletion, it does affect the stability of the plasmid vector. ColE1-related plasmids are not stably maintained in *recBC sbcB* hosts (Ream *et al.*, 1978; Basset & Kushner, 1984; Cohen & Clark, 1986); other replicons such as low copy-number plasmids and *E. coli* minichromosomes are similarly affected (Silberstein & Cohen, 1987). The absence of exonucleases V and I in *recBC sbcB* cells allows the synthesis of linear plasmid multimers by a rolling-circle mechanism (Cohen & Clark, 1986), and multimerisation of high copy-number plasmids can affect their stable maintenance in *E. coli* (Summers & Sherratt, 1984). Rolling-circle replication is initiated by homologous recombination via the RecF pathway (Basset & Kushner, 1984; Luisi-DeLuca *et al.*, 1989), which operates in *recBC sbcB* cells (see review by Clark & Low, 1988). Mutation of *recF* inactivates the RecF pathway, thereby preventing the establishment of rolling-circle synthesis. Hence the requirement for this mutation by Boissy & Astell (1985) was due to their use of ColE1-derived plasmids. Those groups using bacteriophage vectors (Leach & Stahl, 1983; Nader *et al.*, 1985) had no need of *recF* deficiency to ensure vector stability.

On the other hand, the effect of *recBC* and *sbcB* mutations on instability is most probably a consequence of the enhanced viability of palindrome-bearing vectors in such strains, as will be described later. It is worth noting that certain mutant alleles of *recB* and *recC* may even stimulate excision of Tn*10* and Tn*5* (Lundblad *et al.*, 1984), and that *sbcB* mutations have been isolated which enhance illegitimate recombination (Allgood & Silhavy, 1991).

More recently the *recJ* gene has been implicated in palindrome instability: Ishiura *et al.* (1989) found that propagation of recombinant cosmid DNA libraries in a *recBC sbcB recJ* host prevents *recA*-independent deletions, although they do not specify whether the unstable sequences are palindromic. Doubt has been cast on the significance of this finding by Doherty *et al.* (1993), who failed to preserve palindromes cloned in λ phage by growth in SURE™ (*recB sbcC recJ umuC uvrC*) and SRB (*sbcC recJ umuC uvrC*) strains (Stratagene, LaJolla, California). While the authors advise against the use of *recJ* phage hosts for the propagation of palindromic sequences, it must be remembered that the pleiotropic nature of the mutations in these hosts will, if taken in concert, compromise the growth of any DNA sequence. Furthermore, as the *recJ* gene product is a component of the RecF recombination pathway, its mutation will most probably have a stabilising effect only in a *recBC sbcB* background.

Another aspect of the host genotype that affects the deletion of palindromes is DNA repair: excision of Tn*10* is stimulated by mutations in *ssb* (coding for single-strand binding protein), *mutD* (defective in the exonucleolytic proof-reading subunit of DNA polymerase III and in mismatch repair (Schaaper, 1989)), and *uvrD*, *mutH*, *mutL*, *mutS* and *dam* (Lundblad & Kleckner, 1984), which encode components of the *E. coli* methyl-directed mismatch repair pathway (reviewed by Modrich, 1989). This system may repair the replication errors generated by hairpins emanating from palindromes. The participation of *E. coli* DNA gyrase in palindrome instability has also been suggested: the deletion in *E. coli* of an inverted repeat sequence of 1.1 kb (found in the plasmid pDG1 isolated from *Dictyostelium*) can be prevented by the use of a *gyrA* strain (Saing *et al.*, 1988). Models for DNA gyrase-mediated illegitimate recombination (reviewed by Ikeda, 1990) cannot readily account for such deletions using short direct repeats as endpoints.

Hypotheses to Account for Palindrome-Mediated DNA Instability

The *recA*-independence of palindrome deletion qualifies it for admission into an eclectic company of DNA rearrangements, collectively known as illegitimate recombination (Franklin, 1967; reviewed by Ehrlich, 1989). The name bestowed upon these phenomena is appropriate to their lack of common ancestry; models accounting for such events are seldom universal. In the case of palindrome-mediated DNA instability, two alternative hypotheses have emerged and they are summarised below.

Deletion by Replication Slippage

The mutagenic potential of repeated sequences was realised by Streisinger *et al.* (1966) in a study on frameshift mutations in the bacteriophage T4 lysozyme gene. It was proposed that the local misalignment of the complementary DNA strands during replication or repair could act as the precursor of a frameshift mutation. Although the authors confined their work to sequences consisting of the simple reiteration of a few bases, more complex repeats were subsequently shown to have similar mutagenic properties (Farabaugh *et al.*, 1978; Albertini *et al.*, 1982).

The ability of palindromic sequences to catalyse deletions and frameshifts using direct repeats as endpoints, which has been outlined above, has necessitated the incorporation of palindromes into these slipped mispairing, or "slippage" models. Numerous schemes have been proposed (Ripley, 1982; Ripley & Glickman, 1983), but the variant described by Glickman & Ripley (1984) best encompasses the available data. It proposes that the self-complementarity of palindromes allows the formation of

DNA secondary structures that serve as deletion intermediates; these stem-loop structures juxtapose the direct repeats that act as deletion endpoints. The formation of a hairpin would be facilitated in single-stranded DNA. The stability of this structure would depend on the size of the double-stranded stem and the single-stranded loop, which are dictated by the repeat length and the extent of asymmetry at the palindrome centre respectively. DNA polymerase pausing at the base of the hairpin would allow the transient dissociation of the nascent strand, its reannealing to the distal direct repeat on the template strand, and the resumption of synthesis bypassing the stem-loop structure. This process is illustrated in Fig. 1.3.i, and is conceptually related to copy-choice recombination (Lederberg, 1955) in that the key step involves the switching of the DNA replication machinery from one direct repeat to another on a single-stranded template. Replication slippage has also been associated with the evolution of simple repetitive DNA sequences in higher eukaryotes (review by Levinson & Gutman, 1987; Schlötterer & Tautz, 1992), and human genetic disease (Sinden & Wells, 1992). A form of replication slippage has been invoked to account for gene amplification and the appearance of a tandem array of inverted repeats in mammalian cell lines (Hyrien *et al.*, 1988).

Palindromic and quasi-palindromic sequences have been implicated in other mutations such as deletions of only a few nucleotides, or base-substitutions, which are thought to result from the looping out of mispaired bases in hairpin stems (Ripley, 1982; de Boer & Ripley, 1984; reviewed by Ripley, 1990). Additionally, palindromes in heteroduplex DNA appear to inhibit mismatch repair in yeast (Nag *et al.*, 1989); a high frequency of post-meiotic segregation is thought to result from hairpin formation by palindromic insertions of 14 bp or more (Nag & Petes, 1991). However, the ability of palindromes to focus deletions between direct repeats, in terms of both frequency and endpoint location, encourages the conclusion that replication slippage in the manner described above is the predominant cause of palindrome-mediated instability.

Deletion by Hairpin Cleavage

The alternative hypothesis to account for palindrome deletion involves the specific cleavage of hairpin structures. The extrusion of a cruciform structure from a negatively supercoiled double-stranded DNA palindrome would present a substrate consisting of two opposing hairpins. When cleaved across the base of these hairpins by a conformation-specific nuclease, the cruciform would be converted into a double-strand DNA break. Limited 3' $\rightarrow$ 5' exonucleolytic digestion at this break would expose the complementary strands of the direct repeats that previously flanked the

palindrome. Annealing of these sequences, ligation and repair synthesis would eliminate both the double-strand break and one copy of the direct repeats; the product of this reaction is therefore identical to that generated by replication slippage. The hairpin cleavage model for palindrome deletion is illustrated in Fig. 1.3.ii.

Central to this scheme is the cleavage specificity of the nuclease that cuts the stem-loop structures across the base. Enzymes with such an activity have been identified in a wide variety in organisms: bacteriophage T4 endonuclease VII (Kemper & Garabett, 1981; Mizuuchi *et al.*, 1982a) and bacteriophage T7 endonuclease I (deMassy *et al.*, 1985; de Massy *et al.*, 1987) are well characterised, and are thought to perform a debranching function after phage DNA replication. Endonucleases that cleave four-way DNA junctions have been identified in HeLa cell extracts (Waldman & Liskay, 1988), human placenta (Jeyaseelan & Shanmugam, 1988) and calf thymus (Elborough & West, 1990); a nuclease from *Saccharomyces cerevisiae* with a cruciform-specific activity (West & Körner, 1985) is the subject of great research interest. The *ruvC* gene product of *E. coli* has been shown to act on cruciform structures, and has been implicated in the resolution of Holliday junctions in homologous recombination (Connolly *et al.*, 1991; Iwasaki *et al.*, 1991). Its biological role is reviewed by West & Connolly (1992).

However, the cleavage site-specificity shown by these enzymes effectively disallows their involvement in palindrome-mediated deletion. The activities of both T4 endonuclease VII and T7 endonuclease I are directed by the local nucleotide sequence of the four-way DNA junction, and show strong preference for cleavage at pyrimidine residues (Picksley *et al.*, 1990; Pottmeyer & Kemper, 1991). The cruciform-specific yeast endonuclease and the *E. coli ruvC* gene product both depend on the presence of homologous DNA sequences at the junction point (Parsons & West, 1988; Dunderdale *et al.*, 1991). It is therefore unlikely that the instability of palindromic sequences in *E. coli* is due hairpin cleavage by RuvC.

The continuing absence of an appropriate enzyme activity means that the hairpin cleavage model for deletion formation comes a poor second to the replication slippage model. Moreover, only the slippage model can readily account for the complex interplay of direct and inverted repeats in palindrome-mediated instability. It is nevertheless possible that the nuclease-dependent mode of palindrome deletion operates alongside the replication slippage pathway in *Escherichia coli*.

Figure 1.3

Models to Account for Palindrome Deletion (after DasGupta *et al.*, 1987)

i. Replication Slippage

a. Palindromic sequence in double-stranded DNA flanked by a pair of direct repeats.

b. Palindrome forms a hairpin structure at a single-stranded region of the replication fork.

c. DNA synthesis by the polymerase is impeded by the presence of the hairpin.

d. Polymerase pausing facilitates dissociation of the nascent strand from the proximal direct repeat, and its reannealing to the distal direct repeat.

e. Resumption of synthesis stabilises the mispairing and prevents subsequent copying of the extruded sequence.

f. Second round of replication eliminates the palindromic sequence and one copy of the direct repeats.

ii. Hairpin Cleavage

a. Palindromic sequence in double-stranded DNA flanked by a pair of direct repeats.

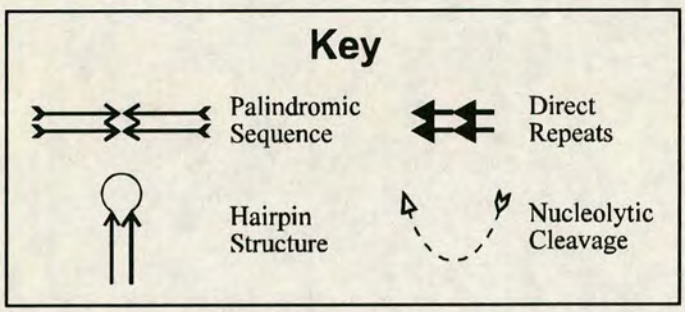
b. Palindrome extrudes to form a cruciform structure.

c. Hairpin-specific nuclease cleaves across the base of the cruciform arms.

d. Product of cleavage is a double-stranded DNA break.

e. Limited 3' → 5' exonucleolytic digestion allows annealing of complementary sequences from the direct repeats.

f. Ligation and repair synthesis restores one copy of the direct repeats.

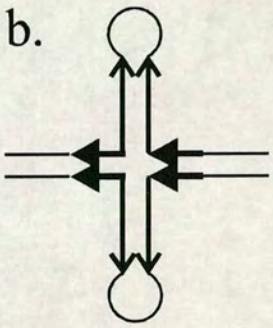


i.

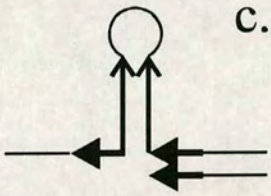
a.

ii.

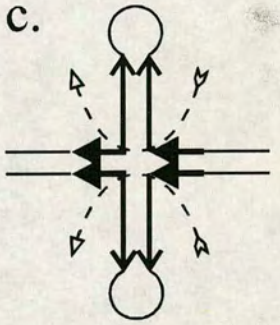
b.



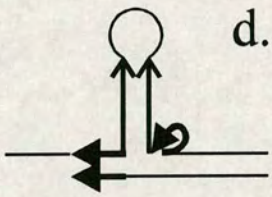
b.



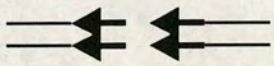
c.



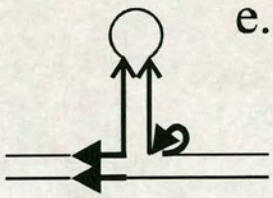
c.



d.

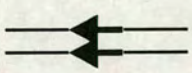


d.



e.

e.



f.

Palindrome-Mediated Replicon Inviability

Extent of Palindrome-Mediated Replicon Inviability

The phenomenon of palindrome-mediated replicon inviability is distinct from the phenomenon of DNA sequence instability described above. A vector containing a palindrome may be viable, but will nevertheless suffer extensive deletions. On the other hand, an inviable replicon may be granted a new lease of life by excision of the offending palindromic sequence.

The reported absence of palindromic sequences in the pioneering use of cosmid vectors by Collins & Hohn (1978) has been followed by numerous accounts of *in vitro* constructs that fail to transform *E. coli*. Plasmid vectors containing inverted dimers of regulatory DNA sequences (Casadaban & Cohen, 1980), palindromic Tn*5* derivatives (Collins, 1981) and inverted plasmid fragment repeats (Lilley, 1981a), and a head-to-head ligation of near-full length pBR322 (Mizuuchi *et al.*, 1982b) all fail to grow in wild-type *E. coli*. λ vectors are similarly crippled by resident palindromes (Leach & Stahl, 1983), while growth of phage ϕX174 is particularly sensitive to the presence of inverted repeats (Müller & Turnage, 1986).

This lethality of palindromic DNA has been put to a constructive use in positive-selection plasmid cloning vectors (Hagan & Warren, 1982; Elhai & Wolk, 1988; Altenbuchner *et al.*, 1992). The principle of this method is simple: the plasmid vector harbours long inverted repeats flanking symmetrical polylinker sequences, which are separated by non-palindromic DNA. Removal of this central stuffer fragment by restriction endonuclease cleavage, and religation of the remaining plasmid restores the palindromic symmetry and thereby prevents the transformation of wild-type *E. coli* by this construct. This inviability may be prevented by the replacement of stuffer fragment with the DNA sequence of interest; positive selection for the recombinant plasmid is therefore effected by the inviability of the non-recombinant vector.

Determinants of Palindrome-Mediated Replicon Inviability

As the phenomena of palindrome-mediated DNA instability and replicon inviability overlap considerably, much of the work on palindrome deletion outlined above could equally well be discussed within the context of palindrome lethality. To avoid repetition the following summary will therefore focus only on determinants of vector inviability that differ from those covered previously in the analysis of sequence

instability. Owing to the importance of the host genotype to palindrome-mediated lethality, the effects of *recB*, *recC*, *sbcB*, and in particular, *sbcC* mutations warrant a separate exposition. This will be followed by a résumé of tenable hypotheses for palindrome inviability.

Palindrome Length

The palindrome length parameters for replicon inviability are broadly similar to those that have been determined for sequence instability. A palindromic sequence of 260 bp confers inviability on its plasmid vector (Lilley, 1981a), whereas a 147 bp perfect palindrome does not compromise the growth of its carrier replicon in *E. coli* (Bergsma *et al.*, 1982). Enlargement of this sequence by 3034 bp endows it with the same lethality exhibited by longer inverted repeats; this can be circumvented by spontaneous *in vivo* deletions removing the centre of symmetry (Hagan & Warren, 1983). These conclusions were endorsed by the work of Yoshimura *et al.* (1986): plasmids containing palindromes of 190 bp or less can multiply in wild-type *E. coli*, whereas vectors with inverted repeats of 368 bp or more are inviable. However, palindromic sequences of a sub-lethal length (146 bp and 147 bp) can elicit intermediate effects on their carrier replicon, such as a reduction in plasmid copy number (Warren & Green, 1985).

While such results may give the impression of a 200 bp size limit for perfect palindromes, care must be taken to differentiate between instability and inviability as the cause of intolerance of longer sequences. For example, the deletion of large palindromes at a high frequency could easily be mistaken for inviability, while the genuine lethality of such sequences would prevent an analysis of their instability.

Asymmetry at the Palindrome Centre

That central interruptions to the inverted symmetry can overcome the inviability associated with such sequences was identified by Collins *et al.* (1982) and Mizuuchi *et al.* (1982b), and then characterised extensively by Warren & Green (1985). The latter authors determined an insertion of 57 bp to be the minimum required for alleviation of lethality. In this respect, replicon inviability differs from palindrome-mediated instability, for the sequence with a 57 bp insert remained highly unstable; only a 150 bp interruption generated stable constructs.

A recent study on the viability of λ phage containing palindromes with differing lengths of central asymmetry has indicated that cruciform extrusion *in vivo* may be responsible for the lethality of inverted repeats (Chalker *et al.*, 1993). It was

found that perfect palindromes confer a more deleterious phenotype than those with a central interruption, and that the severity of inviability declines with the increasing length of asymmetry in the range of 0 to 27 bp. The hierarchy in plating of these λ phage is independent of the host genotype. These results implicate a centre-dependent reaction in the phenotypic effects of palindromic sequences. Cruciform extrusion *in vitro* is thought to be initiated by the denaturation of the central 8 – 10 bp (Murchie & Lilley, 1987; Courey & Wang, 1988; Zheng & Sinden, 1988); changes to the palindrome centre in the form of asymmetric inserts would therefore alter the rate of cruciform extrusion *in vivo*. The lethality to the λ phage of such an unusual DNA secondary structure would account for the results outlined above.

DNA Replication

The replication of a palindromic DNA sequence is required not only for its deletion, but also for the inviability of its carrier replicon. This was demonstrated by Leach & Lindsey (1986): when an *E. coli rec*$^+$ host is infected with a λ phage bearing a lethal 530 bp palindrome, a reduction in the level of intracellular supercoiled λ DNA is observed. However, the reduction is not seen in a *rec*$^+$ strain lysogenic for bacteriophage λ; replication of the superinfecting (palindrome-containing) λ DNA is repressed in this host. The importance of DNA synthesis to replicon inviability was confirmed by Shurvinton *et al.* (1987). It was found that λ phage bearing palindromes of 8.4 kb and 530 bp could be recovered intact following the infection of a *rec*$^+$ *dnaBts* host in which DNA replication had been blocked by growth at the restrictive temperature of 42°C. On the other hand, when DNA synthesis is allowed by growth at 37°C, the palindrome phage that have undergone replication are recovered in greatly reduced numbers, when compared to a control phage without a palindrome. The authors propose that DNA replication drives the extrusion of cruciforms from palindromic sequences *in vivo*, thus forming substrates for conformation-specific nucleases.

That cleavage of the carrier replicon is responsible for palindrome-mediated inviability has been questioned by Lindsey & Leach (1989). The authors used demethylation to study the fate of λ DNA strands introduced into an *E. coli rec*$^+$ *dam* strain (defective for the Dam methylase). They found that a decrease in the yield of palindrome-containing molecules with two newly synthesised (unmethylated) strands can occur without any concomitant loss of replicated (hemimethylated) molecules containing input (methylated) strands. This implies that the inviability of the palindrome phage is not a consequence of DNA cleavage, but of its slow replication.

Carrier Replicon

The type of vector housing the palindromic sequence can determine the extent of inviability, but not directly. The relocation of a palindrome from one plasmid to another does not alter the lethality of that sequence to its carrier replicon (Hagan & Warren, 1983; Yoshimura *et al.*, 1986). The propagation of long perfect palindromes in λ phages is fraught with similar problems (Leach & Stahl, 1983). However, when used in conjunction with a *recBC sbcBC* host that is permissive for large inverted repeats, λ vectors circumvent the problems of plasmid maintenance associated with this *E. coli* genotype.

The degree of replicon inviability is affected only when a single-stranded vector is used to propagate long palindromes. Müller & Turnage (1986) were unable to clone palindromic repeats of longer than 48 bp in bacteriophage ϕX174, and Leach *et al.* (1987) found that insertion of a 110 bp palindrome into bacteriophage M13 rendered the vector inviable. Both phage generate single-stranded molecules during their life cycles, which may account for their heightened sensitivity to the effects of long DNA palindromes.

Effect of *sbcC* and Other Host Mutations on Palindrome-Mediated Replicon Inviability

Alleviation of Palindrome-Mediated Replicon Inviability

The *recB sbcB* mutations were initially found by Collins *et al.* (1982) to lead to a 5 – 10 fold increase in the stability of long palindromes in *E. coli*. However, Leach & Stahl (1983) later showed that the primary effect of this genotype is to permit the growth of palindrome-bearing vectors; the reduced instability is therefore a consequence of the increased viability of the replicon. Moreover, while the latter authors found propagation of a phage λ derivative carrying a perfect palindrome of 3.2 kb to be possible in an *E. coli recBC sbcB* host, this long sequence remains unstable and undergoes symmetrical deletions to give inverted repeats of 600 ±100 bp total length. A similar observation was made by Shurvinton *et al.* (1987): a λ phage with a palindrome of 8400 ±100 bp can plate on an *E. coli recBC sbcB* host, but continues to suffer from sequence instability. Deletion derivatives were isolated carrying palindromes of 700 ±100 bp, suggesting that this is the maximum size limit for stable inverted repeats in the absence of replicon inviability.

The use of *recBC sbcB* mutant *E. coli* hosts has allowed the propagation of complete human and *Drosophila* genomic libraries in λ vectors (Wyman *et al.*, 1985;

Wyman *et al.*, 1986). The alleviation of replicon inviability in these strains is not due to interference of the palindrome with DNA packaging or the recombinational activities of RecBCD. Leach & Lindsey (1986) have shown that a reduction in the level of intracellular supercoiled λ DNA can be seen following infection of a *rec*$^+$ host by a palindrome-bearing phage. This reduction also occurs in a *recA* strain, but not in one carrying mutations in *recBC sbcB*. The absence of general recombination in a *recA* host (Clark & Margulies, 1965) prevents the packaging of *red gam* λ phage DNA. Bacteriophage lacking the *gam* gene product (γ) fail to inactivate the RecBCD enzyme (exonuclease V) and are therefore unable to progress from the θ (Cairns) to the σ (rolling-circle) mode of replication, yielding only circular λ monomers (Enquist & Skalka, 1973). A deficiency in the phage-encoded Red pathway makes the production of packageable multimers dependent on host recombination functions: RecBCD-mediated recombination, which is stimulated by the χ$^+$ (*chi*$^+$) hotspot, catalyses the formation of viable substrates for DNA packaging (see reviews by Smith, 1983; and Smith, 1988b). As this pathway is inoperative in the absence of RecA, the loss of palindrome-bearing phage DNA cannot be attributed to RecBCD-promoted recombination. Likewise, the hypothesis that inviability occurs at the level of dimerisation or packaging must also be ruled out, for σ replication is permitted in a *recBC* host.

Similarly, the activation of the RecF recombination pathway in a *recBC sbcB* host has been discounted as a reason for the alleviation of palindrome-mediated inviability. Leach *et al.* (1987) found that *recBC sbcB* hosts remain permissive for λ phages carrying a 560 bp palindrome when additional mutations in *recA*, *recF*, *recJ*, *recL*, *recN*, *ruv* and *lexA* are introduced. The products of these genes constitute the RecF pathway of recombination which operates in a *recBC sbcB* background (Clark & Low, 1988). The recent discovery that λ encodes an analogue of the *E. coli* RecF pathway genes *recO recR* and *recF* (Sawitzke & Stahl, 1992) does not challenge the conclusions of Leach *et al.* (1987). Bacteriophage λ continues to require the RecA and RecJ proteins for recombination in a *recBC sbcB* cell, but mutations in neither the *recA* nor *recJ* genes affect the viability of palindrome-bearing phage in a *recBC sbcB* strain.

SbcCD Protein

Mutation of the *E. coli sbcC* gene has been found to be primarily responsible for the alleviation of palindrome-mediated inviability (Chalker *et al.*, 1988). The *sbcC* mutation had been discovered by Lloyd & Buckman (1985) in *recBC sbcB* strains,

where acts as a cosuppressor (with *sbcB*) of the defects in DNA repair, recombination and cell viability associated with the *recBC* genotype; strains with mutations in *recBC* and *sbcB*, but not in *sbcC*, grow very poorly. On the other hand, mutation of *sbcC* alone has no effect on recombination in a cell that is otherwise rec^+, and does not suppress the recombination deficiency of *recB* or *recC* mutants. Additionally, plasmid vectors are relatively stable in *sbcC* single mutants (Chalker *et al.*, 1988). This is in stark contrast to *recBC sbcB(C)* mutations, which interfere with plasmid maintenance in *E. coli*.

Chalker *et al.* (1988) used *sbcC* mutant hosts to overcome the inviability of a λ *red gam* derivative containing a 571 bp near-perfect palindrome in *E. coli*. It was found that the λ phage formed plaques with high efficiency on strains with an *sbcC* mutation, irrespective of their *recB*, *recC* or *recD* genotype. The only difference seen in hosts with additional mutations in *recB*, *recC* or *recD* was an increased burst size, which is due to the production of multimeric, packageable DNA by rolling-circle replication as outlined above. This difference could be minimised by the introduction of a χ^+ site into the λ vectors bearing the palindrome; enhanced RecBCD-mediated recombination at this χ^+ site increases the pool of dimeric λ DNA substrate for packaging. This plating behaviour can be seen in λ phage without long inverted repeats, indicating that only the *sbcC* mutation is responsible for the alleviation of palindrome-mediated replicon inviability. This was further confirmed by the finding of Chalker *et al.* (1988) that mutation of *sbcC* alone permits as good a recovery of supercoiled palindromic λ DNA molecules as the combination of *recBC sbcB(C)* mutations used previously (Leach & Lindsey, 1986).

An interaction between the *E. coli sbcC* gene and the bacteriophage λ *gam* gene was shown by Kulkarni & Stahl (1989). Expression of the *gam* gene renders the host cell phenotypically $recBC^-$ (Unger *et al.*, 1972) due to a biochemical inhibition of the RecBCD enzyme by the λ γ protein (Unger & Clark, 1972). γ is required for the transition of λ replication from the θ to the σ mode (Greenstein & Skalka, 1975), and this requirement is abolished in a *recBC* mutant host (Enquist & Skalka, 1973). Kulkarni & Stahl (1989) found that while λ *gam* phage carrying long palindromes require the host *sbcC* mutation for plating, gam^+ λ with the same inverted repeats grow on wild-type *E. coli*, *sbcC* mutants and strains deleted for the *recB*, *recC* and *recD* genes. A similar observation had been made earlier that the propagation of inviable clones from a human genomic library is made possible in rec^+ *E. coli* cells by a functional phage γ protein (Wertman *et al.*, 1986). These results suggest that γ inhibits the action of both the RecBCD and SbcC protein. Speculation that SbcC may

mirror some of the activities of RecBCD has consequently arisen. Riley & Anilionis (1978) have proposed that the *E. coli* genome has undergone two duplications, with the result that homologous genes can be found one quarter or half way around the circular chromosome. It is therefore of interest that *sbcC* is located 180° from the *recBCD* genes on the *E. coli* genetic map (Lloyd & Buckman, 1985).

The *sbcC* gene encodes a poorly expressed protein of 118 kDa (kiloDaltons), and DNA sequence analysis has revealed that it lies immediately downstream of an open reading frame (*orf-45*) encoding a polypeptide of 45 kDa (Naom *et al.*, 1989). The overlap of the TGA stop codon for *orf-45* with the ATG start codon for *sbcC*, and the absence of an independent promoter for the latter has suggested the transcriptional and translational coupling of these two genes. Identification of *orf-45* as a cosuppressor of *recBC* which facilitates the propagation of λ phage bearing long palindromes has led to it being renamed *sbcD* (Gibson *et al.*, 1992). Both the sequence data and the functional relationship between *sbcC* and *sbcD* imply the association of their gene products as a multi-subunit enzyme in the style of RecBCD. This somewhat tenuous link was strengthened recently by the discovery that an open reading frame in *Bacillus subtilis* with a high degree of protein sequence homology to the *E. coli sbcD* gene is located downstream of the *addAB* genes, which constitute the *B. subtilis* analogue of RecBCD (Sharples & Lloyd, 1993).

Although SbcC and RecB, and SbcD and RecC show some homology, it is limited to small regions that are not colinear with the overall protein sequence. Far greater similarity to SbcCD is shown by two putative bacteriophage nucleases, encoded by genes *46* and *47* of bacteriophage T4 (gp46 and gp47), and genes *D13* and *D12* of bacteriophage T5 (gpD13 and gpD12) (Leach *et al.*, 1992). While SbcC, gp46 and gpD13 are more closely related than SbcD, gp47 and gpD12, the genes encoding both the phage nucleases show the same genomic organisation as *sbcCD* (Gram & Rüger, 1985; Kaliman *et al.*, 1988). All three proteins are members of the UvrA-related superfamily of nucleotide-binding proteins that includes many membrane transport proteins (Gorbalenya & Koonin, 1990).

The bacteriophage T4 protein encoded by genes *46* and *47* has been studied in some detail, and is essential for the degradation of the bacterial chromosome upon phage infection (Kutter & Wiberg, 1968). Genetic recombination of phage T4 is reduced in *46⁻47⁻* infections (Cunningham & Berger, 1977), and the failure of such mutants to generate gaps from DNA nicks under *lig⁻* conditions (Prashad & Hosoda, 1972) is held responsible this phenomenon.

The *E. coli* SbcCD protein may therefore constitute a similar nuclease that can generate single-stranded DNA from double-stranded DNA. In the absence of exonuclease V and exonuclease I such an activity would result in persistent single-stranded molecules, which could provide the SOS inducing signal responsible for the filamentation seen in *recBC sbcB* (*sbcC*$^+$) mutants (Lloyd & Buckman, 1985). Such a role for SbcCD is seemingly irreconcilable with the observation of Lindsey & Leach (1989) that slow replication of palindromic DNA causes the inviability of its replicon. However, if the action of the putative SbcCD nuclease were restricted to only one of products of DNA replication (e.g. by degrading only the daughter chromosome containing the lagging-strand template) then no net synthesis would take place, and yet one of the parental strands would remain intact.

Hypotheses to Account for Palindrome-Mediated Replicon Inviability

In general, hypotheses accounting for the inviability of palindrome-carrying vectors have invoked a cruciform or hairpin structure as a target for conformation-specific nucleases. A different proposal was made by Lilley (1981a): the release of negative supercoiling by cruciform extrusion would have deleterious consequences for processes such as transcription, which depend in part on the topological status of the DNA. However, the homeostatic regulation of the *E. coli* DNA gyrase genes (Menzel & Gellert, 1983) would ensure that such a perturbation of supercoiling could not persist, and consequently the effects on replicon viability would be minimal.

The cleavage of an extruded cruciform has been central to most tenable hypotheses for palindrome inviability. The presupposed existence of an enzyme with such an activity, which would normally be responsible for the resolution of recombination intermediates (Holliday junctions), has lent weight to these ideas. Although such an endonuclease has now been identified in *E. coli* as the product of the *ruvC* gene (Connoly *et al.*, 1991; Iwasaki *et al.*, 1991), it has not met with all the functional requirements of palindrome-mediated inviability. The target site specificity of four-way junction-specific nucleases, which has been outlined in the context of palindrome-mediated sequence instability, discourages their involvement in the inviability of palindrome-bearing replicons. In the case of the RuvC protein, cleavage of a cruciform structure depends on homologous DNA sequences at the junction point (Dunderdale *et al.*, 1991). This condition does not appear to be a requisite of palindrome-mediated inviability, and therefore undermines the role of RuvC and other Holliday junction resolvases in this phenomenon.

Furthermore, models of palindrome-mediated inviability that include cruciform cleavage steps resembling Holliday junction resolution are incompatible with the results of Lindsey & Leach (1989), which show that at least one of the parental (input) strands of a palindrome-containing λ phage infection of *E. coli* remains intact, despite the inviability of the replicon. A modified hypothesis, which proposes that the nuclease activity is confined to only one of the products of replication, would encompass these results and retain the central tenet of hairpin-specific cleavage. The role of nuclease would then fall upon the SbcCD protein, mutation of which is a requisite for the propagation of long DNA palindromes. While a nucleolytic activity of SbcCD on stem-loop structures has yet to be shown *in vitro*, its amino-acid sequence homology to the products of bacteriophage T4 genes *46* and *47*, and bacteriophage T5 genes *D13* and *D12* suggests that SbcCD may be similarly responsible for DNA degradation *in vivo*.

A scheme to account for the available data outlined above has been proposed by D.R.F. Leach (unpublished), and is illustrated in Fig. 1.4. The long palindrome is extruded as a cruciform, and assumes one of the isomeric forms of the X-structure known to exist *in vitro*, in which pairs of arms achieve colinearity by terminal stacking of the helices (Duckett *et al.*, 1988). DNA synthesis, emanating from an origin of replication to one side of the cruciform, will therefore encounter one of the stem-loop structures as no more than a continuation of the double helix. The DNA polymerase will therefore denature that hairpin, which is shown here on the leading strand, ensuring its accurate replication. On the other hand, the stem-loop on the lagging strand will persist owing to the failure of DNA polymerase to dissociate the self-complementary repeats that constitute the hairpin.

This obstinate hairpin then becomes the target for the SbcCD nuclease; speculation on its mode of action will not be entertained here, given the paucity of knowledge about this protein. The ensuing double-stranded DNA break will then be subject to exonucleolytic degradation, precipitating the demise of that product of replication. This chain of events will therefore result in no net DNA synthesis, and yet provides a role for SbcCD as a conformation-specific nuclease.

The function of SbcCD *in vivo* may be to prevent error-prone DNA synthesis by the removal of mutagenic substrates such as hairpins; the products of replication slippage are therefore destroyed before they become unrepealable. The small (≤40 bp) and imperfect palindromes found in wild-type *E. coli* would not be expected to extrude as cruciforms on subsequent attempts at accurate replication, whereas the

unnaturally long inverted repeats constructed *in vitro* would elicit the degradative response of SbcCD repeatedly.

The major caveat that should be included with a presentation of this hypothesis for replicon inviability concerns the strand-specificity of both palindrome lethality and deletion. A prediction issued by this model is that mutagenesis by replication slippage would predominate on the lagging stand, for this is what the action of SbcCD seeks to avoid. However, as discussed earlier, the evidence for preferential deletion of palindromic sequences on the lagging strand of synthesis is ambiguous, and probably dependent on the utilisation of DNA polymerase I (Pol I). Current models for DNA replication (see review by Marians, 1992) involve only a limited participation of Pol I, restricting the influence of this enzyme on slippage mechanisms. It is nevertheless possible that only lethal DNA palindromes (longer than those used by Weston-Hafer & Berg (1991)) cannot be replicated by Pol III in the lagging strand. Pol I would then be employed in a form of postreplicative gap repair (Friedberg, 1985) by discontinuous synthesis of the palindromic DNA. Preferential mutagenesis on the lagging strand arising from replication errors by Pol I (Trinh & Sinden, 1991) would therefore be dependent on the presence of long DNA palindromes, which cause replicon inviability in the presence of SbcCD. This hypothesis has yet to be tested.

Concluding Remarks

Palindromic DNA sequences are relatively ubiquitous: their inverted symmetry makes them ideal targets for dimeric DNA-binding proteins, and their ability to form hairpin or cruciform structures provides an additional opportunity for interaction with various aspects of cellular metabolism. However, a disparity in terms of palindrome length and perfection exists between eukaryotic and prokaryotic organisms. The absence of long inverted repeats in the latter is probably a consequence of the instability of these sequences and the inviability they confer upon their carrier replicon. These two components of the behaviour of long DNA palindromes have been thoroughly researched in *E. coli*, and the results of these studies attribute responsibility for deletion and lethality to the formation of cruciforms. These unusual DNA secondary structures will be the subject of the next section.

Figure 1.4

Model to Account for Palindrome-Mediated Replicon Inviability (D. Leach, unpublished)

a. Palindromic sequence in double-stranded DNA.

b. Palindrome extrudes to form a cruciform structure.

c. DNA synthesis initiates at origin of replication.

d. Replication fork proceeds towards cruciform.

e. Leading strand synthesis dissociates the strands of the upper hairpin.

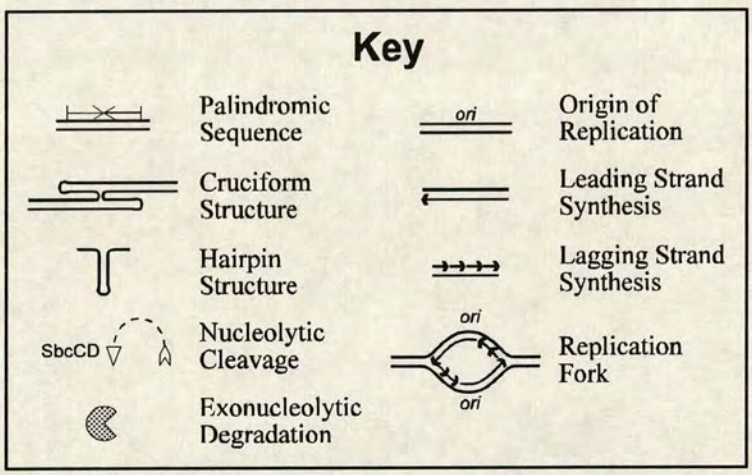
f. Palindrome faithfully replicated by leading strand synthesis, while persistent lower hairpin prevents lagging strand synthesis of palindromic DNA.

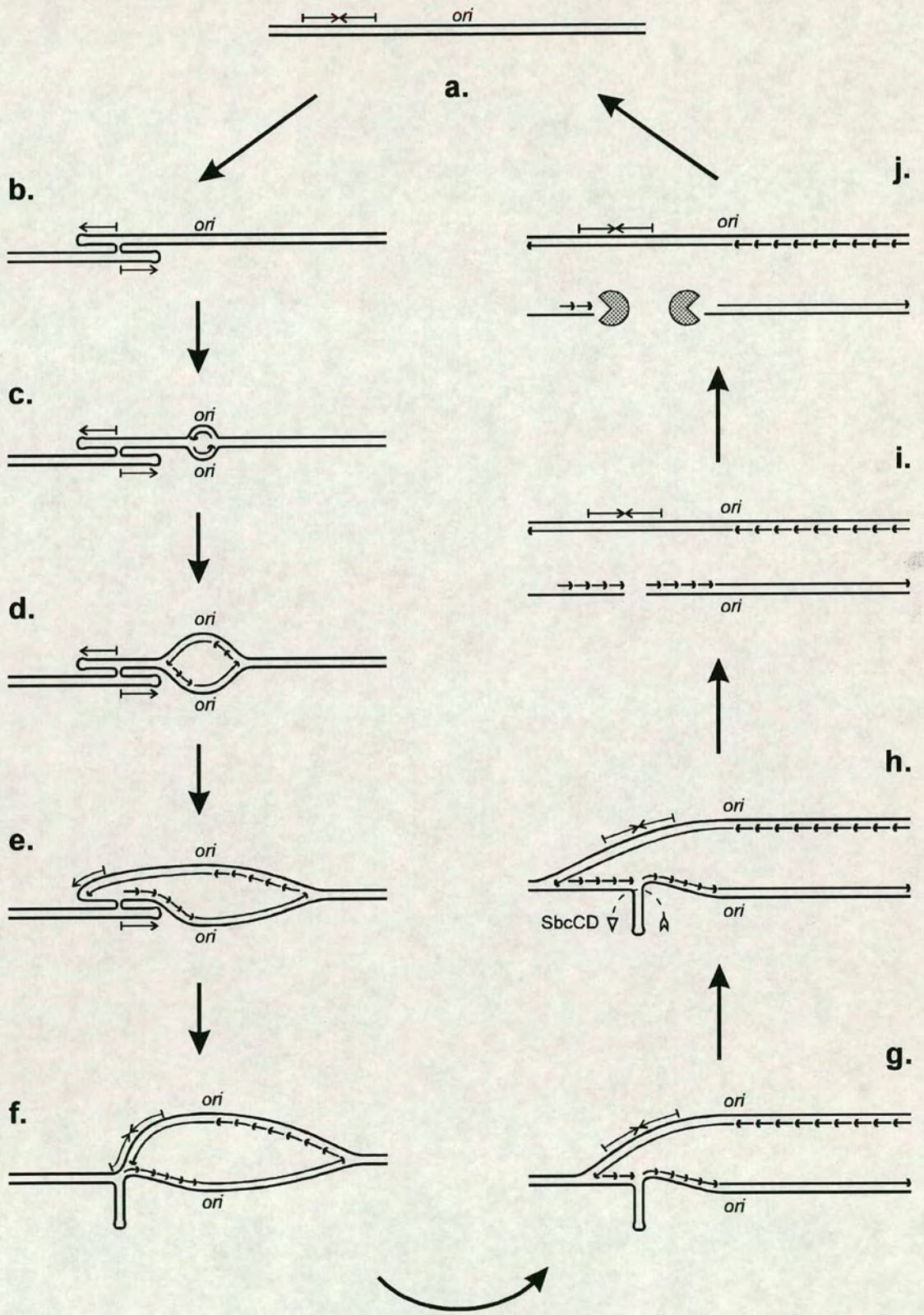
g. Replication proceeds on both strands, but hairpin remains on template for lagging strand synthesis.

h. Hairpin structure is cleaved or degraded by the action of SbcCD.

i. Product of cleavage is a double-stranded DNA break.

j. Exonucleolytic digestion degrades the cleaved replication product, resulting in no net DNA synthesis.




ori
a.
b.
c.
d.
e.
f.
g.
h.
i.
j.
SbcCD

SECTION 2

CRUCIFORM DNA

The extrusion of cruciform structures has long been a theoretical possibility (Platt, 1955), and the capacity of the interstrand base pairs of a palindromic DNA sequence to rearrange into the intrastrand base pairs of two opposing hairpins, as shown in Fig. 1.1, has continued to fascinate researchers (Gierer, 1966; and others). The advent of techniques for detecting DNA conformational microheterogeneity heralded a thorough investigation of cruciform extrusion: the determinants of this secondary structure transition, its thermodynamics, and its kinetics have been studied with exemplary rigour.

However, despite this extensive knowledge of cruciform extrusion, surprisingly little is known about the occurrence of these structures *in vivo*. The fragility of cruciform structures precludes their detection after the relatively harsh conditions of DNA purification; large-scale denaturation of DNA by phenol-chloroform extraction has been shown to introduce structural artefacts such as DNA hairpins (Hyrien, 1989). An *in situ* assay is therefore required, but the methods that had been used successfully *in vitro* are on the whole inappropriate to this approach. Consequently, much of the evidence for and against the existence of cruciform DNA *in vivo* is fraught with conjecture.

This section of the introduction will begin with a short review of the techniques that have made the study of cruciform structures possible, culminating in the method described in this thesis. The results of these studies will then be discussed with respect to the determinants and mechanisms of cruciform extrusion, followed by a short description of cruciform structure. Lastly, a résumé of both real and postulated instances of cruciforms *in vivo* will be given.

Detection of Cruciform DNA

With a few exceptions, methods for the detection of cruciforms have relied upon the interaction of a diagnostic probe with either one of the two characteristic features of this structure: the four-way DNA junction at the base of the cruciform, or the single-stranded DNA loop at the end of the arms. These techniques are summarised below; an overview of DNA structure probes is given by Lilley (1992a).

Electron Microscopy

The detection of hairpins or cruciforms by electron microscopy is often used to identify inverted repeats or palindromes when conventional methods of DNA sequence analysis are inappropriate. This method was widely used before the

introduction of DNA sequencing, such as in the characterisation λ*dv* plasmids (Chow *et al.*, 1974). Heteroduplexes prepared using these phage λ deletion derivatives and wild-type λ DNA formed visible hairpin structures, implying that the λ*dv* strands contained inverted repeats. Electron microscopy of heteroduplex molecules has also been used to demonstrate the presence of palindromic sequences in human genomic recombinant libraries prepared in λ phage (Wyman *et al.*, 1985; Donlon *et al.*, 1986).

More commonly, the rapid self-association of inverted repeats in denatured DNA has been used to prepare hairpin and cruciform structures for electron microscopy. Wilson & Thomas (1974) used this technique to provide evidence for palindromes in eukaryotic DNA, and subsequent examinations of genomic inverted repeats (Schmid *et al.*, 1975; Wesley, 1975; Karrer & Gall, 1976; Nader *et al.*, 1985) have continued to employ this method. As discussed later, electron microscopic analysis of figure-eight molecules and their derivatives, which contain Holliday junctions resembling cruciforms, has proved invaluable in studies of genetic recombination (Potter & Dressler, 1979; and others). However, the use of electron microscopy in the investigation of cruciform extrusion has been relatively limited (Gellert *et al.*, 1979; Borst *et al.*, 1984; Mizuuchi *et al.*, 1982b; Yoshimura *et al.*, 1986). The requirement for cruciform arms of a readily visible size, necessitating the use of long palindromes that cannot be readily propagated in *E. coli*, and the availability of more quantitative techniques have restricted its application.

Gel Electrophoresis

The identification of cruciform structures in circular double-stranded DNA by gel electrophoresis relies primarily upon the reduction of negative supercoiling that is concomitant with extrusion. The theoretical aspects of DNA supercoiling will be discussed in greater detail later, but a working understanding of gel electrophoresis requires the following assumptions. The DNA double helix isolated from living cells is underwound, typically by about one turn in twenty. The torsional tension arising from this negative supercoiling is freely partitioned between two types of structural deformation: twist, which is the rotation of strands about the duplex axis, and writhe, which measures writhing of the duplex axis in space. This geometric distortion of the supercoiled molecule leads to a change in its average shape, which in turn results in altered electrophoretic mobility. A relaxed DNA circle displays no writhing and is therefore largely planar; it will consequently migrate more slowly than a supercoiled molecule with an interwound structure. These two species of circular DNA which differ only in their degree of supercoiling are called topoisomers.

Negative supercoiling favours the extrusion of cruciforms: the energetic cost associated with the loss of base pairing in the single-stranded loop is offset by the relaxation of the underwound DNA. Because the two strands of palindromic DNA are no longer interwound when in the cruciform conformation, the helical turns they previously displayed are redistributed in the remaining part of the molecule; this reduces the twist component of supercoiling, making extrusion energetically favourable. It is therefore possible to separate cruciform DNA from its palindromic progenitor by gel electrophoresis, as the extruded species will appear more relaxed and will migrate more slowly in an agarose gel. Gellert *et al.* (1979) found that cruciform circular DNA has a mobility close to that of nicked (relaxed) circular DNA. This finding has been used by other researchers to measure the kinetics and thermodynamics of extrusion (Mizuuchi *et al.*, 1982b; Lilley & Hallam, 1984).

Two-Dimensional Gel Electrophoresis

The resolving power of one-dimensional agarose gel electrophoresis for topoisomers is impressive: a ladder of species is observed, each differing from the other by one turn of the duplex. However, the range over which migration is a function of topology is limited; in particular, the comigration of relaxed molecules with the cruciform species can often obscure this supercoiling-dependent structural transition. Two-dimensional gel electrophoresis increases the sensitivity of this method considerably, and is initially no different to the one-dimensional variety: a complete set of topoisomers is electrophoresed in a single lane of an agarose gel. The gel is then soaked in an intercalating agent, usually chloroquine. Intercalation leads to a local unwinding of the helix (Bauer & Vinograd, 1968), which must compensated by positive writhe; this leads to positive supercoiling of the molecule, the extent of which is directly proportional to the amount of chloroquine added. This will remove all the cruciform structures that were previously present. The gel is then rotated 90° and electrophoresed for a second time, with chloroquine present in the running buffer.

Those topoisomers with the least negative supercoiling (slowest migration in the first dimension) will show the greatest positive supercoiling after intercalation, and will therefore migrate fastest in the second dimension. It is consequently possible to distinguish cruciform and non-cruciform species that comigrate in one-dimensional gel electrophoresis, as extrusion will be reversed by positive supercoiling. The two molecules will migrate as distinct topoisomer spots, for the ex-cruciform, which was previously more negatively supercoiled than the relaxed comigrant, will now be the less positively supercoiled of the two species. This method of detecting cruciforms

was first employed by Lyamichev *et al.* (1983) and Courey & Wang (1983), and has continued to be widely used for research into the determinants of extrusion (Gellert *et al.*, 1983; Panyutin *et al.*, 1984; Greaves *et al.*, 1985; Courey & Wang, 1988). It has been reviewed by Bowater *et al.* (1992).

Comparative Gel Electrophoresis of Kinked DNA

Gel electrophoresis has also been used to investigate the structure of the four-way DNA junction at the base of a cruciform. Gough & Lilley (1985) found that a pseudo-cruciform junction, generated by heteroduplex formation between two palindromes, exhibits extremely anomalous gel electrophoretic mobility. Their results were similar to those obtained by Wu & Crothers (1984) using kinked trypanosome kinetoplast DNA, indicating that the cruciform induces a bend in the linear DNA molecule. The anomalous migration of cruciforms in polyacrylamide gels has facilitated research into effects of divalent cations on this structure (Diekman & Lilley, 1987), and has permitted analysis of the geometry of the four-way DNA junction (Cooper & Hagerman, 1987; Duckett *et al.*, 1988), including the influence of base mismatches on the branch point structure (Duckett & Lilley, 1991). The application of comparative gel electrophoresis to the elucidation of global features of DNA structure is reviewed by Crothers & Drak (1992).

Psoralen Cross-Linking

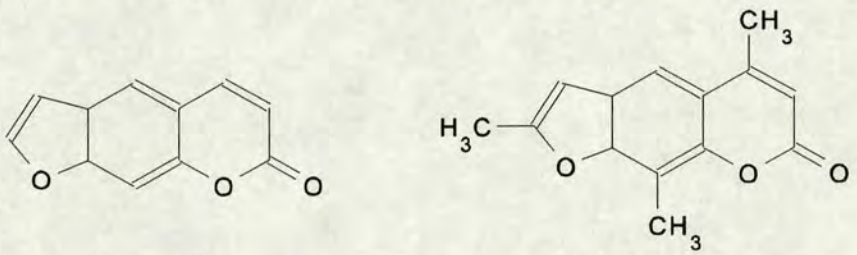
Psoralens are widely used as probes of DNA structure, especially *in vivo*: bacterial and eukaryotic cells are permeable to these chemicals (Fig. 1.5) which can photobind to nucleic acids. Psoralens initially bind noncovalently to DNA by intercalation; this interaction is sufficiently weak to disturb the topology and structure only minimally. Irradiation with near-UV light (320 – 400 nm) photobinds an intercalated psoralen to an adjacent pyrimidine base, forming a monoadduct. When the adjacent base in the opposite strand is also a pyrimidine, interstrand cross-links are formed in a second photochemical reaction.

Psoralen cross-links can be assayed in a number of ways, which rely upon the rapid renaturation of the strands that are covalently held together; hydroxyapatite chromatography (as described earlier) and denaturing gel electrophoresis are examples of such assays. Recently developed techniques using exonuclease digestion or primer extension reactions allow the sites of psoralen photoaddition to be mapped accurately (Ussery *et al.*, 1992). Psoralen photobinding will cross-link the intrastrand base pairing of a cruciform, locking this otherwise unstable structure in the four-way

configuration. The fraction of the DNA that contains cruciform structures can subsequently be estimated by restriction endonuclease digestion at sites that will liberate a full-length palindrome from linear DNA, or half-length hairpins from cruciform DNA (Sinden *et al.*, 1983; Sinden & Pettijohn, 1984). This procedure is described in detail by Ussery *et al.* (1992). Psoralen cross-linking of cruciforms has been employed as a torsionally tuned probe for measuring DNA supercoiling *in vivo*: the level of negative supercoiling can be determined from the fraction of palindromes of a predetermined length existing in the cruciform conformation *in vivo* (Zheng *et al.*, 1991; Sinden & Ussery, 1992). Psoralen cross-linking has also been used to demonstrate the formation of DNA hairpin artefacts during phenol-chloroform extraction of cellular DNA (Hyrien, 1989).

Figure 1.5

Molecular structure of psoralen and 4,5',8-trimethylpsoralen.



Four-Way DNA Junction Resolution

The characteristic structure of the four-way DNA junction (Sigal & Alberts, 1972) invites speculation about its interaction with proteins; in particular, nucleases that resolve such a joint molecule have received much attention. Two such enzymes are endonuclease VII of bacteriophage T4 and endonuclease I of bacteriophage T7; both are used to detect cruciforms. T4 endonuclease VII resolves Holliday junctions (Mizuuchi *et al.*, 1982a) by cleavage of the crossover strands (Mueller *et al.*, 1988), and specifically cleaves cruciforms in supercoiled plasmids (Lilley & Kemper, 1984). Consequently, it is widely used as a probe of cruciform extrusion *in vitro* (Lilley & Hallam, 1984; Gough & Lilley, 1985; Greaves *et al.*, 1985; Lilley, 1985; Naylor *et al.*, 1986; Pontiggia *et al.*, 1993; del Olmo & Pérez-Ortín, 1993).

T7 endonuclease I can also resolve branched DNA substrates and cleave cruciforms (de Massy *et al.*, 1985); like T4 endonuclease VII, it cuts the opposing strands at or near the branch point (de Massy *et al.*, 1987). T7 endonuclease I has

recently been used as an alternative probe of four-way DNA junctions *in vitro* (Murchie & Lilley, 1987; Frappier *et al.*, 1989; Glucksmann *et al.*, 1992). The specific interaction of both T4 endonuclease VII and T7 endonuclease I with Holliday junctions has been used to elucidate the molecular geometry of this structure (Duckett *et al.*, 1988; Bhattacharyya *et al.*, 1991; Duckett *et al.*, 1992).

Anti-Cruciform DNA Monoclonal Antibodies

Cruciform detection in bacterial plasmids is reasonably straightforward, given the small size of the vector. However, the observation of unusual DNA structures in a mammalian genome that is several orders of magnitude larger is a daunting task. One possible solution involves the use of monoclonal antibodies, which have been raised against stable cruciform structures in a heteroduplex DNA molecule (Frappier *et al.*, 1987). The binding of anti-cruciform monoclonal antibodies is specific to such a DNA conformation but is independent of DNA sequence, and has been shown to occur at the four-way junction (Frappier *et al.*, 1989). These antibodies have been used *in situ* to determine the role of cruciform structures in mammalian DNA replication (Zannis-Hadjopoulos *et al.*, 1988), their presence throughout the cell cycle (Ward *et al.*, 1990) and their association with the nuclear matrix (Ward *et al.*, 1991). They have also been used in the anti-cruciform DNA affinity purification of active mammalian origins of replication (Bell *et al.*, 1991). All of these results will be discussed later.

Single-Strand DNA-Specific Nuclease Probes

S1 Nuclease

The single-stranded DNA loop at the tip of the cruciform arms is a diagnostic feature of this unusual structure, and nuclease digestion at this loop is frequently used to assay extrusion (reviewed by Wohlrab, 1992). The most commonly used probe is the single-strand-specific endonuclease S1 from *Aspergillus oryzae*, which has been employed in cruciform loop detection since the identification of palindromes in eukaryotic genomes (Wilson & Thomas, 1974; Engberg *et al.*, 1976). These authors found no evidence of S1 nuclease sensitivity in hairpins formed by self-associated palindromic DNA and concluded that the extent of asymmetry separating the inverted repeats was negligible. However, S1 nuclease digestion was later repeated using artificial palindromic sequences in supercoiled plasmids: cleavage was found to occur specifically at the centre of the inverted repeats, and was dependent on negative supercoiling (Lilley, 1980; Panayotatos & Wells, 1981). The recognition and cleavage

of single-stranded DNA hairpin loops by S1 nuclease has been used to investigate the thermodynamics and kinetics of cruciform extrusion *in vitro* (Lilley, 1981b; Singleton & Wells, 1982; Lilley & Markham, 1983; Lilley & Hallam, 1984; Greaves *et al.*, 1985; Lilley, 1985; Naylor *et al.*, 1986; Sullivan & Lilley, 1986; Murchie & Lilley, 1987; Sullivan & Lilley, 1988; Murchie & Lilley, 1989). S1 digestion has also been used to probe for cruciform DNA in naturally occurring palindromes (Noirot *et al.*, 1990; del Olmo & Pérez-Ortín, 1993), and hairpin structures at sites of protein-DNA interaction (Waga *et al.*, 1990; Pognan & Paoletti, 1992). The major problem associated with S1 nuclease is its acidic pH optimum (pH 4.5 – 5), which restricts its use under physiological conditions.

Other Single-Strand DNA-Specific Nucleases

Other single-strand-specific endonucleases have consequently been used as alternative probes for cruciform DNA. *Bal*31 nuclease cleavage at inverted repeats in supercoiled DNA is coincident with that of S1 nuclease, and has therefore been used to detect hairpin loops (Lilley & Hallam, 1984; Greaves *et al.*, 1985; Naylor *et al.*, 1986; Morales *et al.*, 1990). Similarly, mung bean nuclease has been shown to cleave the single-stranded regions of cruciform structures (Sheflin & Kowalski, 1985) and has been used to determine the structure of DNA hairpin loops (Xodo *et al.*, 1991). Micrococcal nuclease has been used sporadically as a probe of cruciform extrusion (Dingwall *et al.*, 1981; Greaves *et al.*, 1985). T7 endonuclease I, which cleaves four-way DNA junctions as described above, has an additional nucleolytic activity on the single-stranded loops of cruciform structures (Panayotatos & Wells, 1981). This property has occasionally been employed in the detection of cruciforms *in vitro* (Orlowski & Miller, 1991). Arguably the most significant account of cruciform extrusion as measured by T7 endonuclease I loop cleavage was reported by Panayotatos & Fontaine (1987): induction of nuclease synthesis by transcription of a cloned T7 gene *3* leads to specific cleavage at the centre of a DNA palindrome *in vivo*. This result is discussed later, but it is appropriate to note here that uncertainty about the target site specificity of T7 endonuclease I has yet to be resolved.

Single-Strand DNA-Specific Chemical Probes

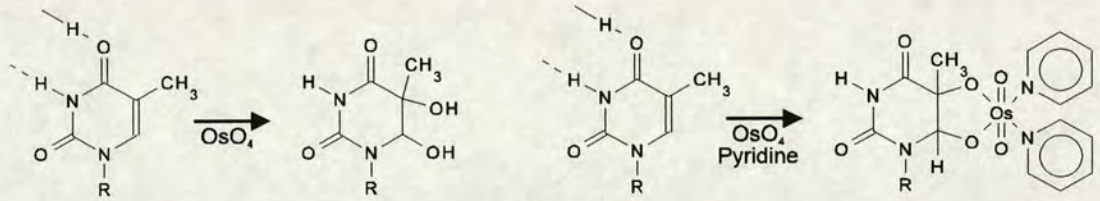
Osmium Tetroxide

Chemical probing using reagents that specifically react with the unpaired bases of the hairpin loop is currently the most popular means of studying cruciform

extrusion *in vitro*. Osmium tetroxide, which is used in electron microscopy as a fixative and staining agent for biological tissues, is arguably the most widespread of these single-strand-specific chemical probes. As shown in Fig. 1.6, it reacts with unpaired pyrimidine bases (preferentially thymine) to yield *cis*-pyrimidine glycol, and in the presence of suitable tertiary amines such as pyridine and 2,2'-bipyridine forms stable diester adducts. This can subsequently be detected in a number of ways: cleavage by S1 nuclease followed by gel electrophoresis is often sufficient to confirm osmium modification, as the modified bases are rendered permanently single-stranded. However analysis at single-nucleotide resolution is most commonly accomplished by use of the Maxam-Gilbert sequencing protocol (reviewed by Maxam & Gilbert, 1980); radiolabelling of the DNA fragment, cleavage at the site of the osmium adduct by piperidine and denaturing polyacrylamide gel electrophoresis enables the detection of the reaction site. The use of osmium tetroxide as a probe of DNA structure *in vitro* is reviewed by Palecek (1992a).

Figure 1.6

Reaction of thymine with osmium tetroxide, and with osmium tetroxide and pyridine.



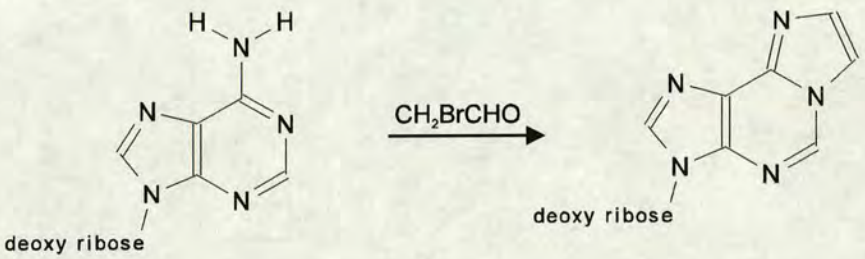
Lilley & Palecek (1984) found that in supercoiled DNA, a site of pronounced hyper-reactivity towards modification by osmium tetroxide mapped to the centre of an inverted repeat. The extrusion of a cruciform structure from this palindrome was held responsible, as osmium tetroxide would react preferentially with unpaired pyrimidines in its hairpin loop. Subsequent research into cruciform extrusion has made frequent use of osmium tetroxide as a probe for single-stranded DNA loops (Lilley & Hallam, 1984; Greaves *et al.*, 1985; Naylor *et al.*, 1986; McClellan & Lilley; 1987; Bowater *et al.*, 1991; Klysik, 1992). The absence of reactivity at the four-way DNA junction of a cruciform has been used to study its conformation and folding (Duckett *et al.*, 1988; Duckett *et al.*, 1990; Duckett & Lilley, 1991). Osmium tetroxide has recently been used to study the formation of cruciform structures by negatively supercoiled $(A\text{-}T)_n$ sequences *in vivo* (McClellan *et al.*, 1990); probing DNA structure by *in situ* chemical modification is reviewed by Palecek (1992b).

Haloacetaldehyde

Cruciform DNA may also be detected by single-strand-specific reactions with haloacetaldehyde: both chloroacetaldehyde and bromoacetaldehyde react with the exocyclic amino groups of unpaired bases, primarily adenine and cytosine, to form etheno derivatives. This reaction permanently disrupts Watson-Crick pairing, and is illustrated in Fig. 1.7. Haloacetaldehyde-modified bases can then be mapped by S1 nuclease digestion or Maxam-Gilbert DNA sequencing, as described above. The major disadvantage of haloacetaldehyde modification is the requirement for acidic reaction conditions, whereas osmium tetroxide can be used at a neutral pH. A recent review of the detection of unusual DNA secondary structures by haloacetaldehyde has been written by Kohwi-Shigematsu & Kohwi (1992).

Figure 1.7

Reaction of unpaired adenine with bromoacetaldehyde to form 1,N^6-ethenoadenine.



Hyper-sensitivity to haloacetaldehyde was mapped by Lilley (1983) to the centre of an inverted repeat in supercoiled DNA; this correlation with other probes of DNA structure has established haloacetaldehyde modification as a common method of cruciform detection *in vitro* (Lilley & Markham, 1983; Kang & Wells, 1985; Bowater *et al.*, 1991). It has also been used *in vivo* to determine cruciform extrusion from natural palindromes (Noirot *et al.*, 1990) and artificial $(A\text{-}T)_n$ sequences (Dayn *et al.*, 1991; Dayn *et al.*, 1992).

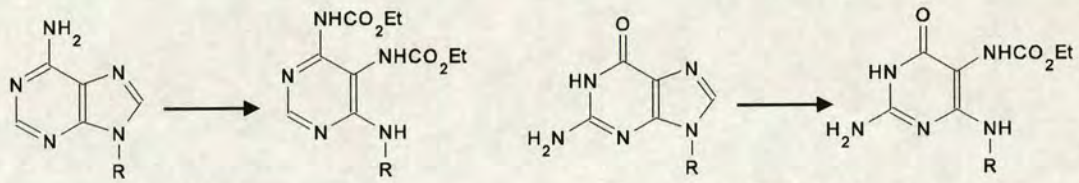
Diethyl Pyrocarbonate

Diethyl pyrocarbonate reacts with the N^6 and N^7 positions of adenine and the N^7 position of guanine when these positions are accessible. It will therefore react with purines only when they are located in single-stranded DNA, or alternatively when they are in the *syn* conformation in a region of Z-DNA (Herr, 1985). The product of the reaction is a ring-opened dicarbethoxylated derivative, as illustrated in Fig. 1.8. As with other single-strand-specific chemical probes, the most common method of

detecting diethyl pyrocarbonate-modified bases is Maxam-Gilbert DNA sequencing. The use of diethyl pyrocarbonate in probing DNA structure is reviewed by Kohwi-Shigematsu & Kohwi (1992).

Figure 1.8

Reaction of diethyl pyrocarbonate with adenine and guanine.



The highly selective chemical modification of cruciform loops by diethyl pyrocarbonate was observed by Furlong & Lilley (1986) and Scholten & Nordheim (1986). Chemical reactivity is restricted to a small region of the palindrome centre, is not found at the four-way DNA junction, and is dependent on negative supercoiling. However, the use of diethyl pyrocarbonate to probe for cruciform extrusion has been somewhat limited (Greaves *et al.*, 1985; Klysik, 1992).

Sodium Bisulphite

In a reaction with sodium bisulphite, cytosine bases in single-stranded DNA are deaminated to deoxyuracil via bisulphite addition across the C5-C6 double bond. The major advantage of probing for cruciform loops with bisulphite lies in the fact that reaction of DNA is not self-propagating, unlike modification by haloacetaldehyde or osmium tetroxide. The method of detecting modified bases (Lilley *et al.*, 1985; Gough *et al.*, 1986) stands in stark contrast to those described above: supercoiled DNA containing a palindrome is reacted with sodium bisulphite and transformed into a repair-deficient *E. coli* strain. The net effect of the base modification is a C $\rightarrow$ T transition, which is subsequently analysed by restriction endonuclease cleavage or DNA sequencing. While sodium bisulphite is arguably the most sensitive single-strand-specific chemical probe of cruciform extrusion, its use is cumbersome by comparison with other methods and has therefore found little application.

Methylase Probes

While enormous progress has been made in the characterisation of unusual DNA structures *in vitro* (reviewed by Wells, 1988), the conformation of DNA *in vivo*

remains largely unknown. Few, if any of the techniques outlined above constitute appropriate methods for the detection of cruciform and other non-B-form DNA structures in living cells. Direct visualisation by electron microscopy or gel electrophoresis necessitates DNA purification, which can produce artefacts in supercoiled molecules. Single-strand-specific chemical and nuclease probes, and psoralen cross-linking are toxic to cells. Furthermore they remove the products of DNA structural transitions, thereby altering the equilibria of reactions such as cruciform extrusion; this predicament is shared by anti-cruciform DNA antibodies.

A new approach has recently emerged that uses DNA methyltransferases as probes of DNA structure; these enzymes will modify only a right-handed Watson-Crick DNA helix. The inhibition of methylation at target sites that are in regions of unusual DNA secondary structure can be detected after DNA purification by cleavage with methylation-sensitive restriction endonucleases. The methylation status is resistant to DNA extraction techniques and thereby serves as a chronicle of DNA-methylase interactions *in vivo*, measuring the steady-state concentrations of methylation-sensitive and resistant structures. The use of methylase probes in living cells is non-invasive, posing little or no viability problems, and the modification is effectively reversible by virtue of DNA replication; the unusual structure must be enduring to inhibit methylation significantly *in vivo*. Moreover, as methylation favours structural transitions such as Z-DNA formation (Singleton *et al.*, 1982) or cruciform extrusion (Murchie & Lilley, 1989) by altering the stability of the DNA helix (Collins & Myers, 1987), the inhibition of methylation would conversely disfavour the formation of unusual DNA secondary structures *in vivo*; self-propagating undermethylation is therefore prohibited. This method of probing DNA structure *in vivo* is reviewed by Zacharias (1992).

Modification by *Hha*I methylase and cleavage by *Hha*I endonuclease are inhibited *in vitro* when the 5' GCGC 3' target site is located in a region of Z-DNA (Vardimon & Rich, 1984; Zacharias *et al.*, 1984). *Eco*RI methylase is also unable to modify its target sites when these are in a left-handed Z-DNA conformation, and has been used to detect the formation of this unusual DNA structure from $(C\text{-}G)_n$ sequences *in vivo* (Jaworski *et al.*, 1987; Jaworski *et al.*, 1988; Jaworski *et al.*, 1989). Triplex H-DNA can similarly inhibit *Eco*RI restriction and modification enzymes *in vitro* (Hanvey *et al.*, 1990), but its detection *in vivo* has been accomplished using the endogenous Dam methylase of *E. coli* (Parniewski *et al.*, 1990). Dam methylase, which is discussed in greater detail below, is unable to modify 5' GATC 3' target sequences within the loop region of a homopurine·homopyrimidine segment capable

of forming H-DNA. Form V DNA, a highly torsionally strained configuration that is obtained by re-annealing unlinked complementary single-stranded DNA circles, has been studied *in vitro* using a variety of sequence-specific methylase probes (Brahmachari *et al.*, 1987; Bagga *et al.*, 1990).

Methylase probes also been used to investigate DNA-protein interactions. The CpG DNA methyltransferase from *Spiroplasma species* has been used to monitor *in vitro* DNA binding of nuclear proteins from HeLa cells (Kochanek *et al.*, 1993). An *in vivo* analysis of DNA-protein interactions in *E. coli* has been performed by Wang & Church (1992). The authors used modification by Dam methylase at GATC target sites to study protein-binding at regulatory DNA sequences. Dam methylase reacts with DNA to produce N^6-methyladenine residues at GATC sequences. This methylation is essential for methyl-directed DNA mismatch correction (reviewed by Modrich, 1989), and is also implicated in the regulation of gene expression, DNA replication, and chromosome segregation (reviews by Marinus, 1987; and Barras & Marinus, 1989). Being endogenous to *E. coli*, Dam methylase represents the zenith of non-invasive probes for DNA structure or protein-DNA interactions in that organism. Nevertheless, it has been used exogenously to analyse the *in vivo* chromatin structure of active genes in *Saccharomyces cerevisiae* (Singh & Klar, 1992).

An omission from the unusual DNA secondary structures listed above is the cruciform. Although Jaworski *et al.* (1987) attempted (and failed) to detect cruciform structures *in vivo* using *Eco*RI methylase, the palindrome they used was relatively short and had an unnatural base sequence ($(TG)_{20}AATT(CA)_{20}$). A systematic analysis of cruciform extrusion from long DNA palindromes *in vivo* using methylase probes has not been reported to date. Such an investigation is presented in this thesis: cruciform extrusion in *E. coli* is measured by the inhibition of Dam methylation at a GATC target site at the centre of a 476 bp perfect palindrome in bacteriophage λ. If the cruciform extrudes during growth of the λ phage in *E. coli*, the GATC sequence will be located in the single-stranded DNA loop and will consequently not be modified by Dam methylase. The ability of cruciform loops to inhibit cleavage by restriction endonucleases (Mizuuchi *et al.*, 1982b) and modification by methyltransferases (Murchie & Lilley, 1989) is well established. After λ DNA purification, the methylation status of the GATC site at the palindrome centre, and control GATC sites outside the palindrome, is examined by cleavage with the methylation-sensitive restriction endonucleases *Dpn*I and *Mbo*I; this technique is described in greater detail in the Materials & Methods and Results chapters.

Extrusion of Cruciform DNA

Determinants of Cruciform Extrusion

The kinetics and energetics of cruciform extrusion (reviewed by Murchie & Lilley, 1992) are primarily determined by the base composition of the palindromic sequence and the level of DNA supercoiling. In this respect, cruciforms resemble other unusual DNA secondary structures such as left-handed Z-DNA and triplex H-DNA: they require negative supercoiling for thermodynamic stability, and the nature and rate of the structural transition is dependent on the DNA sequence.

Thermodynamics — DNA Supercoiling

The two strands of a circular double-stranded DNA molecule are intertwined, and are therefore linked with each other a certain number of times. This is the linking number *Lk*, which for an unconstrained, relaxed molecule of *N* base pairs is given by:

$$Lk = Lk° = N / h \tag{1}$$

where h is the helical repeat; $h°$, the helical repeat of relaxed B-form DNA under physiological conditions, is close to 10.4 bp per turn (Wang, 1979). The linking number may be altered by a nick-ligation activity (topoisomerase) that under or overwinds the DNA; this is termed supercoiling (Vinograd *et al.*, 1965). In the absence of topoisomerase activity, the linking number of closed circular DNA is a topological invariant, and the molecule therefore exists in a number of isomeric forms (topoisomers). The altered linkage of these topoisomers is given relative to the relaxed state ($Lk°$) by means of the linking difference ΔLk:

$$\Delta Lk = Lk - Lk° \tag{2}$$

Most natural DNA molecules are underwound, or negatively supercoiled, so that ΔLk is negative. A more useful term is the specific linking difference, or superhelical density (σ), which is independent of molecular size and is given by:

$$\sigma = \Delta Lk / Lk° \tag{3}$$

Natural DNA molecules are typically underwound by one turn in twenty, which corresponds to a superhelical density of –0.05.

DNA supercoiling demands an input of energy: the free energy of a supercoiled DNA molecule relative to the relaxed state is quadratically related to the linking difference (Bauer & Vinograd, 1970). This free energy of supercoiling (ΔG_s)

has been determined experimentally (Depew & Wang, 1975; Pulleyblank *et al.*, 1975) and is expressed by the function:

$$\Delta G_s = (1050RT / N)\Delta Lk^2 \quad (4)$$

where R is the gas constant (8.314 J K^{-1} mol^{-1}) and T the absolute temperature. The torsional tension arising from negative supercoiling causes a deformation of the molecule which is given by:

$$\Delta Lk = \Delta Tw + Wr \quad (5)$$

where Tw is the helical twist of the DNA strands about their common axis and Wr is the writhing of the helix axis in three-dimensional space. The energy of supercoiling is freely partitioned between these geometric components.

Cruciform extrusion converts interstrand base pairs to intrastrand base pairs, thereby causing a local unwinding of the DNA (Platt, 1955). The twist change (ΔTw_c) resulting from extrusion by a palindromic sequence of n bases is given by:

$$\Delta Tw_c = n / h \quad (6)$$

By redistributing the interwinding of the double-stranded palindrome to the remainder of the molecule, the negative twist change compensates for the linkage deficiency; the DNA thus becomes partially relaxed. The negative free energy change associated with the global relaxation of the molecule can be calculated from Eq. (4) and is offset against the free energy of cruciform extrusion (ΔG_c). ΔG_c arises from the cost of forming two hairpin loops and a four-way DNA junction, and has been determined experimentally to be 17 – 18 kcal mol^{-1} (Courey & Wang, 1983; Gellert *et al.*, 1983; Lilley & Hallam, 1984), although this depends partly on the base composition of the palindrome. The role of DNA sequence in determining cruciform extrusion is discussed later.

Owing to the quadratic relationship between the free energy of supercoiling and the linking difference, cruciform structures will be stable when the level of negative supercoiling is greater than some critical value. Theoretical calculations of this threshold level of supercoiling have relied on thermodynamic (Hsieh & Wang, 1975) and mechanical (Benham, 1982) analysis, and statistical-mechanical theory (Vologodskii *et al.*, 1979; Benham, 1982; Vologodskii & Frank-Kamenetskii, 1982; Vologodskii & Frank-Kamenetskii, 1983; Katsura *et al.*, 1993). The results, which indicate that a superhelical density of around $\sigma \lesssim -0.05$ is needed for cruciform extrusion, have been confirmed by *in vitro* experiments. The requirement for negative DNA supercoiling has been established using a wide variety of cruciforms (Gellert *et*

al., 1979; Lilley, 1980; Lilley, 1981b; Panayotatos & Wells, 1981; Mizuuchi *et al.*, 1982). More detailed studies have shown a dramatic rise in the rate of extrusion over a narrow range of superhelical density (between $\sigma = -0.04$ and $\sigma = -0.06$), above which the cruciform is the stable species (Singleton & Wells, 1982; Courey & Wang, 1983; Gellert *et al.*, 1983; Lilley & Hallam, 1984; Naylor *et al.*, 1986; Courey & Wang, 1988). The effect of supercoiling on DNA structure is reviewed by Frank-Kamenetskii (1990).

While it is well established that circular DNA extracted from bacterial cells has a linking deficit of about one turn in twenty ($\sigma = -0.05$), it is doubtful whether this reflects the available level of negative supercoiling *in vivo*. In particular, DNA complexed with the *E. coli* HU protein is tightly wound, its periodicity being reduced from 10.5 to 8.5 bp per helical turn (Broyles & Pettijohn, 1986). The writhing change that results from nucleoprotein condensation is partially offset by this overwinding, with the net effect that HU restrains only about half the DNA helical tension (reviewed by Pettijohn & Hodges-Garcia, 1990). Estimates of available supercoiling *in vivo* using cruciform extrusion by $(A\text{-}T)_n$ sequences (Greaves *et al.*, 1985; Lilley, 1986) and site-specific recombination in phage λ integration (Bliska & Cozzarelli, 1987) have yielded values of $\sigma = -0.025$, which is in good agreement with the superhelical density calculated for HU-DNA complexes.

$(A\text{-}T)_n$ sequences present no kinetic barrier to cruciform extrusion *in vitro* (this is discussed later), and have therefore been used, in conjunction with *in situ* chemical probing, to measure DNA supercoiling *in vivo*. It has been shown that superhelical torsion in cellular DNA responds directly to genetic and environmental factors such as osmotic shock (McClellan *et al.*, 1990; Dayn *et al.*, 1991), and varies between different sites (Zheng *et al.*, 1991). In particular, more cruciforms are found upstream than downstream of divergent promoters, verifying the twin supercoiling domain model of Liu & Wang (1987); DNA upstream from two divergent transcripts will be negatively supercoiled by transcription, whereas positive supercoils will be introduced downstream. While the estimates of *in vivo* supercoiling provided by these studies have varied between $\sigma = -0.025$ and $\sigma = -0.045$, they have been consistently lower than the values obtained from extracted plasmid DNA. This is of particular relevance to the thermodynamics of cruciform extrusion, which is quadratically related to the superhelical density. The formation of cruciform structures *in vivo* by palindromes of a normal base composition has consequently been questioned.

Kinetics — DNA Sequence, Ionic Conditions and Temperature

Cruciform extrusion *in vivo* has also been called into question as a result of kinetic studies, which have shown that this structural transition is extremely slow. The half-time of extrusion is typically in the range of several minutes to hours, though this depends on the palindromic sequence and length, the degree of supercoiling and the temperature. Although a supercoiled palindrome of several thousand base pairs has an extrusion half-time of three minutes at 35°C (Mizuuchi *et al.*, 1982), shorter sequences (such as a 68 bp palindrome) more typically take about one hour at 37°C to show 50% cruciform formation (Courey & Wang, 1983). Gellert *et al.* (1983) found that even at 50°C, a 48 bp palindrome has an extrusion half-time of several hours; the rate at lower temperatures is even slower. While extrusion rates can vary for different inverted repeats (Sinden & Pettijohn, 1984), the reaction is generally very slow. Such observations have also been made by Panyutin *et al.* (1984), and confirm the theoretical predictions of Vologodskii & Frank-Kamenetskii (1983): the extrusion half-time has a sharp maximum at the superhelix density which corresponds to the equilibrium transition point between the cruciform structure and the regular double helix. While increased levels of supercoiling may result in faster rates of extrusion and ionic conditions may alter the kinetics considerably (discussed later), temperature has arguably the most significant effect.

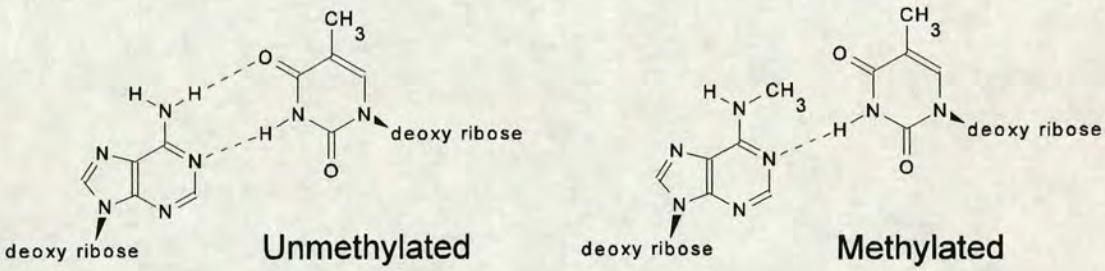
Gellert *et al.* (1983) found the rate of extrusion to decrease by a factor of 10 for every 10°C decrease in temperature, from which the authors calculate an activation energy of 50 kcal mol^{-1}. Similar observations were made by Courey & Wang (1983), who found it necessary to heat a 68 bp palindrome at 65°C to obtain an equilibrium mixture of cruciform and non-cruciform structures. It was also found that the effect of temperature on the kinetics of extrusion is mimicked by agents that decrease the melting temperature of DNA, such as formamide or phenol. Similarly, the effect of supercoiling on the rate of cruciform formation (Courey & Wang, 1988) can also be interpreted in terms of DNA melting; negative supercoiling destabilises the B-form helix (Majumdar & Thakur, 1985; Benham, 1992; Sen *et al.*, 1992).

The effect of helix stability on the rate of cruciform formation is greatest at the centre of the palindrome. In a study of extrusion kinetics using a concerted series of inverted repeats, Murchie & Lilley (1987) found that changes to the central 8 – 10 nucleotides profoundly affected the extrusion rate (by up to 2000-fold), whereas mutations further from the centre altered the rate to a smaller extent. Similar results were obtained by Courey & Wang (1988): replacing the sequence AATT at the centre

of a 68 bp perfect palindrome with the sequence CCCGGG decreases the rate of cruciform formation by a factor of at least 100. The importance of base pairing and base stacking interactions at the centre of symmetry to the kinetics of extrusion was confirmed by Zheng & Sinden (1988). Two series of 96 bp palindromes were constructed with differing centre and arm sequences; in every case inverted repeats with 8 – 10 AT base pairs in the centre were kinetically much more active in cruciform formation than those with 8 – 10 GC base pairs in the centre. Modification of a central GGATCC sequence by DNA methylation may also affect extrusion kinetics. Methylation at N^6-adenine enhances the rate of cruciform formation fourfold by lowering the stability of AT base pairs (Fig. 1.9), whereas C5-cytosine methylation reduces the extrusion rate between 1.7 and 2.7-fold by strengthening GC base pairs (Murchie & Lilley, 1989).

Figure 1.9

Destabilisation of AT base pairs by N^6-methyladenine formation.



In all these studies, a reasonable correlation between the extrusion half-times ($t_{½}$) and the calculated average dinucleotide stability constants (T_{MN}) (Gotoh & Tagashira, 1981) of the central 10 bp of the palindromes can be observed. An empirical relationship has been derived by Murchie & Lilley (1989):

$$T_{MN} = 48.1 + 4 \ln t_{½} \qquad (7)$$

As mentioned above, ionic conditions may alter the kinetics of extrusion considerably. Palindromic sequences of an average base composition (such as those used in the studies outlined above), which appear to require central DNA melting for cruciform formation, exhibit a salt-dependent mode of extrusion. Gellert *et al.* (1983) found that the maximal rate of cruciform formation occurs between 30 mM and 40 mM NaCl, although subsequent research has shown the optimal salt concentration depends partly on the central sequence of the palindrome (Murchie & Lilley, 1987). A systematic analysis of the influence of cation size and charge on salt-dependent cruciform extrusion has been performed by Sullivan & Lilley (1987). It was found that

monovalent group Ia cations most effectively promote extrusion at 50 – 60 mM, while divalent group IIa and selected transition metal ions (notably manganese) are effective over a wide range of concentrations, extending down to 200 μM. A marked correlation between the rate of extrusion and ionic radius was observed, with larger ions giving faster rates of cruciform formation. The authors interpret this as specific ion binding by the transition state of the extrusion process, for which they propose a partially extruded proto-cruciform. Phosphate-phosphate repulsion at this four-way junction creates an anionic "cavity" that is stabilised by cation binding. Palindromes extruding via this salt-dependent pathway are called "S-type" sequences.

Ionic conditions may also affect cruciform extrusion from a different class of palindromic sequences. Termed "C-type" sequences, they are typified by the ColE1 inverted repeat (Oka *et al.*, 1979). Extrusion by this palindrome is suppressed by salt (Lilley, 1985), and shows a very high temperature dependence and a large Arrhenius activation energy (180 kcal mol^{-1}); the extrusion reaction must therefore proceed by a highly co-operative mechanism. Although the ColE1 inverted repeat has a normal base composition, it is flanked by blocks of (A+T)-rich DNA. These flanking sequences confer unusual C-type extrusion kinetics upon the ColE1 inverted repeat; the sequence of the inverted repeat itself appears to have little or no influence (Sullivan & Lilley, 1986). The sequence requirements (Sullivan *et al.*, 1988) and thermodynamics (Schaeffer *et al.*, 1989) of this contextual influence on cruciform extrusion have been analysed. (A+T)-rich sequences from a wide variety of sources can confer C-type kinetics on a *cis* inverted repeat by long range structural communication, and have been termed "C-type inducing sequences". The insertion of DNA between the inducing and palindromic sequences can alter the kinetics of extrusion, depending on the length and (G+C) content of the fragment; these are termed "transmitting sequences". The sequence context of a palindrome can therefore have a profound effect on the rate of extrusion.

Some C-type inducing sequences, such as an $(A\text{-}T)_{34}$ fragment from a *Xenopus* globin gene, may themselves form cruciforms (Greaves *et al.*, 1985). Extrusion by this sequence has a low energy of formation (13.8 kcal mol^{-1}), exhibits no detectable kinetic barrier, and occurs even at low temperatures. $(A\text{-}T)_n$ and other C-type inducing sequences have been examined using structural probes and statistical mechanical models (McClellan & Lilley, 1987; Bowater *et al*, 1991). At low ionic strengths, (A+T)-rich sequences in negatively supercoiled DNA undergo co-operative helix-coil transitions to form denatured regions, the extent of which is a function of temperature and superhelix density. The facile large-scale opening of (A+T)-rich

sequences is the result of stable DNA unwinding, as opposed to transient DNA "breathing" (Kowalski *et al.*, 1988), and may also occur in positively supercoiled DNA (McClellan & Lilley, 1991). Propagation of the base pair opening into palindromic sequences will facilitate the transition to a cruciform structure. A reduction in temperature or an increase in ionic strength changes the properties of (A+T)-rich sequences by stabilising the DNA helix; induction of C-type cruciform extrusion from adjacent palindromes is thereby reduced. Furthermore, low concentrations of distamycin (which specifically stabilises (A+T)-rich DNA) lead to quasi-S-type extrusion from C-type constructs, while helix-destabilising solvents (dimethyl formamide and formamide) confer C-type properties on normally S-type molecules (Sullivan & Lilley, 1988). C-type extrusion and base pair opening in supercoiled DNA have been reviewed by Lilley (1988).

Mechanisms of Cruciform Extrusion

The contrasting kinetic properties of C-type and S-type cruciforms have led to the proposal of two mechanistic pathways for extrusion. They have been reviewed by Murchie *et al.* (1992) and are illustrated in Fig. 1.10. S-type extrusion (Fig. 1.10.i) is initiated by the melting of a small (~10 bp) region at the centre of a palindrome in negatively supercoiled DNA. Intrastrand base pairing by this DNA bubble results in a protocruciform stem-loop structure. This transition state is stabilised by cation binding. Supercoiling-driven branch migration leads to the extrusion of the fully-formed cruciform, which is a thermodynamically stable species. Conversely, C-type extrusion (Fig. 1.10.ii) is initiated by melting of the (A+T)-rich inducing sequences flanking the palindrome; this is facilitated by low ionic strength. Propagation of this unwinding results in the co-operative large-scale opening of the palindromic DNA. This step is characterised by a high activation energy. Intrastrand base pairing of the entire palindrome leads to the closure of this opened region and formation of the extruded cruciform in a single step.

Implications for the possibility of cruciform extrusion *in vivo* are disheartening. It is questionable, on account of DNA binding by protein HU, whether the available level of negative supercoiling is sufficient for cruciform formation. Moreover, extrusion may be kinetically forbidden under physiological conditions, as the relaxation of DNA during replication would prevent cruciform structures from persisting. Because the characteristic half-time of extrusion exceeds the generation time of *E. coli*, the cruciform may not be present in a significant fraction of the DNA, despite its potential thermodynamic stability (Courey & Wang, 1983).

Figure 1.10

Mechanistic Schemes for Cruciform Extrusion (after Lilley, 1985)

i. S-type Cruciform Extrusion

a. Palindromic sequence in double-stranded DNA.

b. Local DNA melting of ~10 bp at the palindrome centre.

c. Intrastrand base pairing and formation of protocruciform transition state in the presence of salt.

d. Branch migration to generate fully extruded cruciform structure.

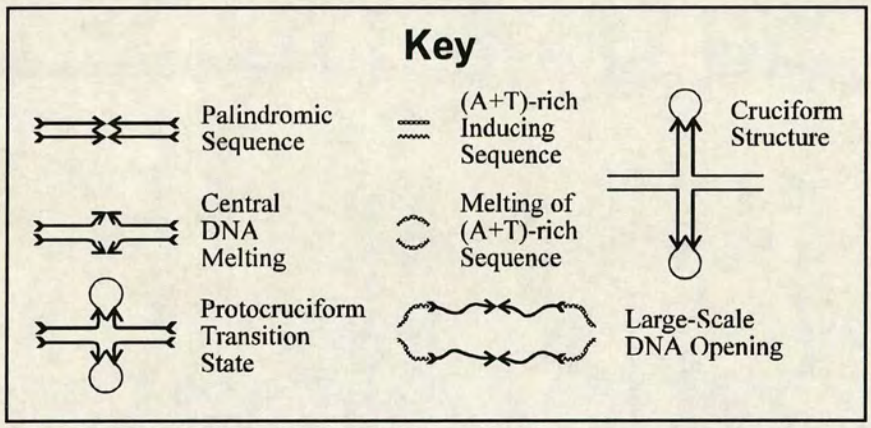
ii. C-type Cruciform Extrusion

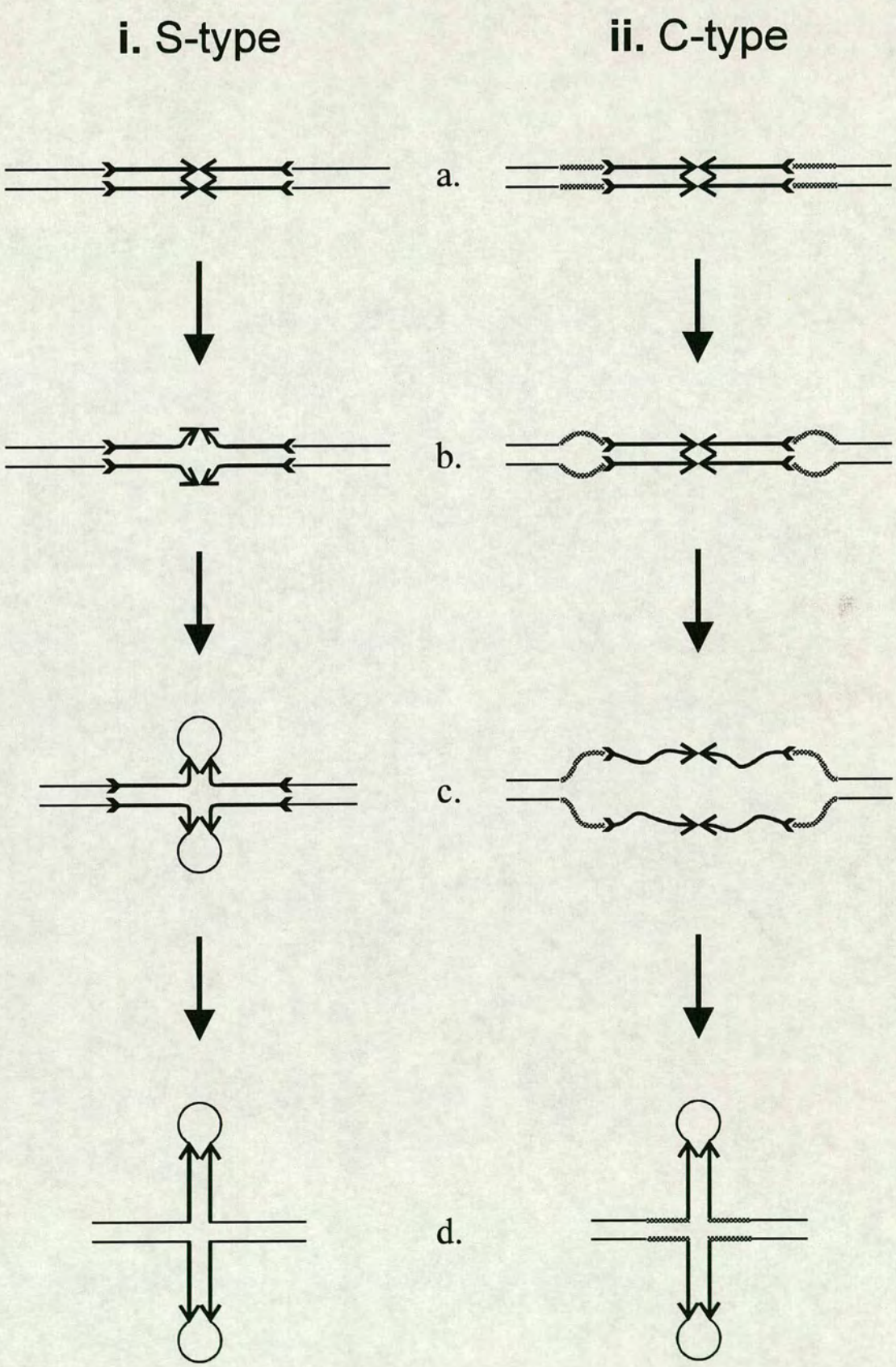
a. Palindromic sequence flanked by (A+T)-rich C-type inducing sequences.

b. (A+T)-rich DNA initiates thermal helix opening in the absence of salt.

c. Propagation of unwinding leads to large-scale opening of DNA encompassing the palindrome.

d. Intrastrand base pairing of entire palindrome generates the fully extruded cruciform in a single step.




i. S-type
ii. C-type
a.
b.
c.
d.

Structure of Cruciform DNA

Cruciform structures are comprised of three components: the four-way junction, the two stems or arms, and the two single-stranded loops. The stems are unremarkable apart from their topological isolation; DNA strand interwinding in the cruciform arms is unaffected by changes in the superhelix density of the remaining molecule. On the other hand, the structure of the hairpin loops and four-way DNA junction are sufficiently distinct from the normal B-form helix to merit a short discussion.

DNA Loops

The DNA loops of a cruciform structure are conventionally represented by four unpaired bases. This is unsurprising, given the evidence for four-membered loops provided by high-resolution chemical probing using sodium bisulphite (Gough *et al.*, 1986). However, NMR (nuclear magnetic resonance) studies (review by Van de Ven & Hilbers, 1988) have shown that in addition to loops of four bases (Kallick & Wemmer, 1991), three-membered loops (Boulard *et al.*, 1991) and even two-membered loops (Rinkel & Tinoco, 1991) can be formed. Two-residue DNA loops have also been observed using gel electrophoresis, UV spectroscopy and nuclease cleavage (Xodo *et al.*, 1988; Howard *et al.*, 1991; Xodo *et al.*, 1991). The conformational feasibility of a hairpin with two purines in the loop has been verified by molecular modelling (Raghunathan *et al.*, 1991).

The thermal stability of a DNA loop is primarily determined by its size, with two-base and three-base loops showing the greatest stability (Germann *et al.*, 1990; Xodo *et al.*, 1991; Rentzeperis *et al.*, 1993). Furthermore, the influence of base composition on the thermal stability of three-membered loops (Germann *et al.*, 1990) corresponds to that observed in four-membered loops (Senior *et al.*, 1988), where the order of stability is:

$$T_4 \text{ loop} > C_4 \text{ loop} > G_4 \text{ loop} > A_4 \text{ loop}.$$

The base sequence may also affect the loop folding in DNA hairpins, thereby determining the size of this single-stranded region. NMR studies of hairpin loops formed by an inverted repeat fragment with a central TTTA region have demonstrated Hoogsteen base pairing between the first and fourth residue of this sequence, thereby producing a two-membered loop (Blommers *et al.*, 1989). A closer inspection of this structure has revealed extensive base stacking interactions and a sharp turn of the

sugar-phosphate backbone in the two-base loop (Blommers *et al.*, 1991). Furthermore, enzymatic studies using *Eco*RI restriction endonuclease cleavage have shown that hairpin loops affect the conformation of stem residues close to them; this effect is greatest for small loops (Germann *et al.*, 1990). Theoretical predictions of DNA hairpin loop structures have similarly revealed complex base interactions (Erie *et al.*, 1993).

It has emerged from these studies that the canonical four-base loops of cruciform structures are an oversimplification. The sequence-dependent formation of two-base or three-base loops can have profound consequences for extrusion. The increased thermal stability of small DNA loops will reduce both the kinetic barrier and thermodynamic cost of cruciform formation.

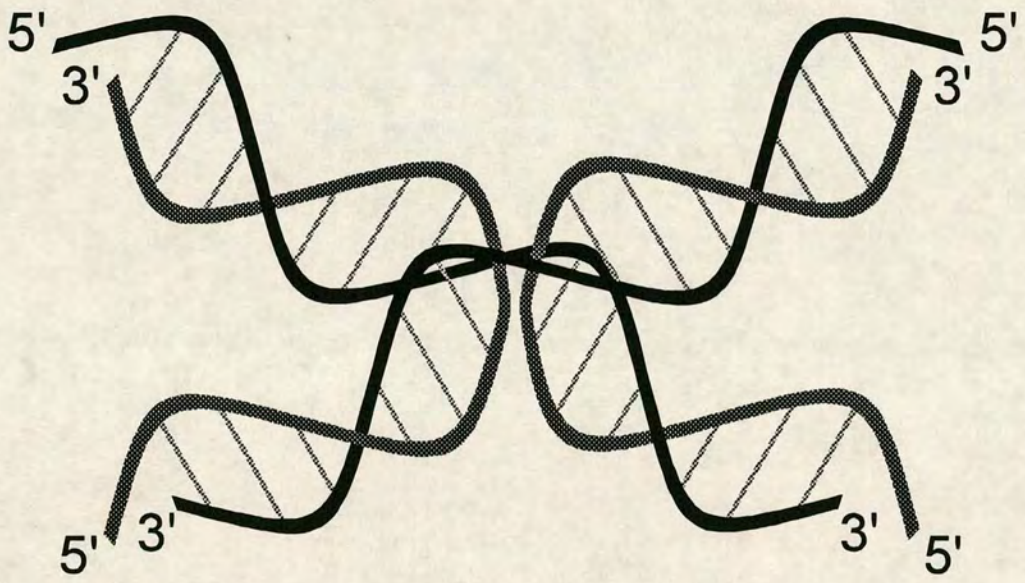
Four-Way DNA Junction

While the stereochemical feasibility of a four-way DNA junction was established in the early years of molecular biology (Sigal & Alberts, 1972), the molecular geometry of this structure has only been determined recently. Stable DNA junctions have been constructed from DNA fragments with sequence symmetry constraints, generating pseudocruciform structures that are unable to undergo branch migration (Bell & Byers, 1979; Kallenbach *et al.*, 1983). Gel electrophoresis of such constructs revealed a considerable retardation in mobility that is indicative of DNA bending at the four-way junction (Gough & Lilley, 1985). This anomalous gel migration is dependent on the concentration of cations (Diekmann & Lilley, 1987; Duckett *et al.*, 1990), and is due to the folding of the DNA junction into an X-shaped structure (Cooper & Hagerman, 1987; Duckett *et al.*, 1988). Electric birefringence decay (Cooper & Hagerman, 1989) and fluorescence energy transfer (Murchie *et al.*, 1989; Clegg *et al.*, 1992) measurements have confirmed and extended this deduction, showing that the four-way DNA junction is a right-handed cross of antiparallel molecules. Stereochemical considerations suggest that the formation of such a structure will minimise unfavourable steric and electrostatic interactions by permitting helix-helix stacking, and is likely to be the most energetically favourable conformation (von Kitzing *et al.*, 1990; Lu *et al.*, 1992). Base pairing at the site of branching has been demonstrated by NMR analysis (Wemmer *et al.*, 1985; Chen *et al.*, 1991) and chemical probing (Gough *et al.*, 1986). Similarly, chemical probing (Churchill *et al.*, 1988), DNase I cleavage (Lu *et al.*, 1989) and restriction endonuclease cleavage (Murchie *et al.*, 1991) have confirmed that the DNA junction is a two-fold symmetric complex whose four arms form two helical domains of pairwise coaxial stacking.

Figure 1.11

Schematic Representation of the Stacked X-Structure for the Four-way DNA Junction (after Murchie *et al.*, 1989)

The model shows a two-fold symmetric structure with pairwise stacking of arms to generate quasi-continuous coaxial helices. The continuous strands (shown in black) have an antiparallel alignment and are accommodated in the major grooves of the opposing helix; this folding is optimal if the helices display a right-handed rotation of approximately 60°. The exchanging strands (shown in dark grey) pass between the two stacked pairs at the point of strand exchange. The two sides of the structure are not equivalent: on the upper side of the model the four base pairs at the point of strand exchange all present minor groove edges, while on the lower side the corresponding major groove edges are presented.



Detailed DNase I probing experiments have revealed protected regions that are consistent with this model (Murchie *et al.*, 1990). The continuous strands of the stacked helices show a displacement of DNase I protection away from the junction towards the 3' side of the strand. The protected region of the backbone is located in the major groove of the opposed helix, avoiding steric clash. This arrangement is favoured if the X-structure forms a right-handed cross with an antiparallel alignment of the continuous strands at an angle of 60°, as shown in Fig. 1.11. The two sides of the stacked X-structure formed by DNA helices at 120° are dissimilar, as the four base pairs at the point of strand exchange are oriented in the same direction. On one side of the junction only the major groove edges are presented, while on the other side only the minor groove edges are presented. The structural non-equivalence is exploited by enzymes that resolve four-way DNA junctions: T4 endonuclease VII specifically cleaves the crossover (exchanging) strands of a DNA junction (Duckett *et al.*, 1988; Mueller *et al.*, 1988) by exclusive interaction with the minor groove edges of two DNA helices at an angle of 120° (Bhattacharyya *et al.*, 1991). The interaction of four-way DNA junctions with enzymes is discussed by Duckett *et al.* (1992).

Cruciform DNA *in Vivo*

Documented examples of cruciforms *in vivo* are rare, owing to the difficulty in applying the methods outlined earlier to living cells. Where cruciforms are found, they often coincide with sites of protein-DNA interactions as predicted by Gierer (1966).

Genetic Recombination

The most widely recognised example of a four-way DNA junction *in vivo* is the Holliday junction. This structure was proposed by Whitehouse (1963) and Holliday (1964) as an intermediate in fungal recombination. Its structural feasibility was subsequently verified by molecular modelling (Sigal & Alberts, 1972) and hydrodynamic calculations (Meselson, 1972). While the Holliday junction is not genuine example of cruciform DNA, the two structures are analogous in that they share the feature of a four-way DNA junction. Although some recombination models have attempted to incorporate both structures (by proposing the initiation of strand exchange at cruciform loops) (Sobell, 1972; Wagner & Radman, 1975), such ideas have become disfavoured. Nevertheless, the common feature of the four-way DNA junction implies that Holliday junction resolvases will also cleave cruciform structures, which may in turn account for the behaviour of palindromic DNA discussed earlier.

Figure 1.12

Schematic Representation of the Holliday Model for Genetic Recombination (after Potter & Dressler, 1976)

a. Two homologous DNA double helices are aligned.

b. In each helix, equivalent strands are nicked at homologous positions.

c. The nicked strands dissociate from their complementary strands.

d. The free ends of the nicked strands associate with the complementary strands in the homologous DNA helices.

e. Ligation of the DNA ends stabilises the reciprocal single-strand exchange, generating a Holliday junction.

f. Continuing reciprocal strand exchange allows movement of the crossover point (branch migration), forming heteroduplex DNA.

g. Alternative view of the Holliday junction.

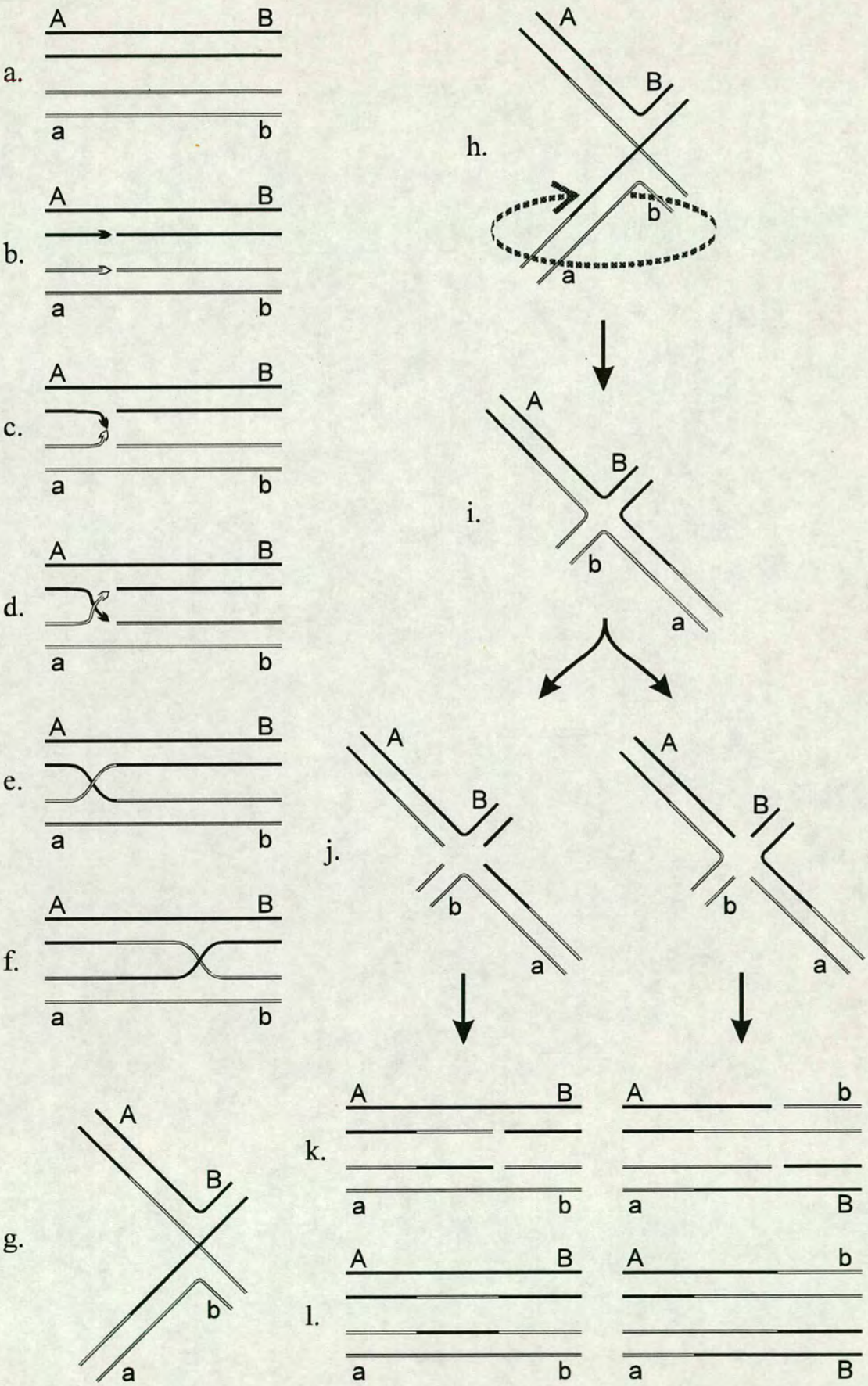
h. Sigal-Alberts isomerisation of the Holliday junction.

i. The Holliday junction shown as a four-way DNA junction.

j. Resolution of the Holliday junction: cleavage of the crossover strands (left branch) or non-crossover strands (right branch).

k. Products of Holliday junction resolution shown as linear DNA molecules: patch (left branch) and splice (right branch) recombination products.

l. Ligation of DNA molecules generates non-recombinant (left branch) and recombinant (right branch) products.


a.
A
B
a
b
b.
A
B
a
b
c.
A
B
a
b
d.
A
B
a
b
e.
A
B
a
b
f.
A
B
a
b
g.
A
B
b
a
h.
A
B
b
a
i.
A
B
b
a
j.
A
B
b
a
A
B
b
a
k.
A
B
a
b
A
b
a
B
l.
A
B
a
b
A
b
a
B

A generalised model for homologous recombination, which is based on the scheme proposed by Holliday (1964), is shown in Fig. 1.12. The Holliday junction is formed in step e. and undergoes Sigal-Alberts isomerisation in step h., resulting in a four-way DNA junction (step i.). The resolution of this structure in step j. by cutting of the crossover strands (left branch) or non-crossover strands (right branch) leads respectively to non-recombinant (patch) and recombinant (splice) products (step l.).

The Holliday model has been periodically updated to reflect genetic data from fungal recombination. The variation proposed by Meselson & Radding (1975) involves initiation of recombination by a single-strand (asymmetric) transfer, which may, after isomerisation, become a two-strand (symmetric) transfer. It has become widely accepted, especially for homologous recombination promoted by χ sites and the RecBCD enzyme of *E. coli* (Smith & Stahl, 1985). An alternative model involves the initiation of recombination at double-strand breaks (Resnick, 1976; Orr-Weaver *et al.*, 1981; Szostak *et al.*, 1983). These double-strand breaks may be enlarged to form double-strand gaps flanked by 3' single-stranded tails. One 3' end can invade an intact homologous duplex, displacing a D-loop. Enlargement of the D-loop by repair synthesis, followed by repair synthesis of the second gapped strand generates two Holliday junctions, which may be resolved to generate crossover or non-crossover products. Although the double-strand-break repair model was proposed to account for genetic recombination in yeast, it has also been used to explain Red-mediated recombination of phage λ (Thaler *et al.*, 1987) and RecBCD-promoted homologous recombination at χ sites in *E. coli* (Rosenberg & Hastings, 1991). The role of double-strand breaks in the recombination of phage λ and yeast has been reviewed by Thaler & Stahl (1988).

A discussion of the mechanisms of genetic recombination is beyond the scope of this introduction; a number of commendable reviews have been published (e.g. Smith, 1988a; West, 1992). However, it is noteworthy that alternative hypotheses for recombination generally differ only in the mechanisms of initiation. The Holliday junction, and its resolution by enzymatic cleavage, has remained a central and unifying feature of these divergent models. The potential for cruciform structures to subvert this stage of recombination encourages a short discussion of Holliday junctions and their cleavage by resolvase enzymes.

Holliday Junctions

Physical evidence for the existence of Holliday junctions *in vivo* has been provided by electron microscopy of phage S13 and ϕX174 replicative form DNA

isolated from recombination-proficient *E. coli* (Doniger *et al.*, 1973; Benbow *et al.*, 1975; Thompson *et al.*, 1975). Four-way DNA junctions formed by recombination between plasmid molecules have also been observed by electron microscopy, both in DNA isolated from *E. coli* (Potter & Dressler, 1976) and in fused structures produced by an *in vitro* recombination system (Potter & Dressler, 1978). The generation of Holliday junctions in *E. coli* is dependent on a functional *recA* gene product, as discussed by Potter & Dressler (1979). RecA protein can also promote the formation of four-way DNA junctions *in vitro* (DasGupta *et al.*, 1981; West *et al.*, 1983).

Holliday junctions have been detected in *S. cerevisiae*: recombination intermediates in 2-μm plasmid DNA preparations from meiotic yeast cells were isolated by restriction endonuclease cleavage (Bell & Byers, 1979). The resultant four-way DNA junctions are resistant to branch migration, and can therefore be prepared by gel electrophoresis and observed by electron microscopy. Psoralen cross-linked chromosomal DNA isolated from pachytene-arrested yeast cells is similarly protected from extensive branch migration, and was shown by electron microscopy to contain branched molecules resembling Holliday junctions (Bell & Byers, 1983).

The detection of ephemeral reaction intermediates, such as the Holliday junction, is often made possible by halting the progress of the reaction. This may be brought about by trapping DNA junctions, such as in the work of Bell & Byers (1979; 1983) outlined above, or by the use of recombination mutations that prevent resolution. The latter method was used by Broker & Lehman (1971) in a study of phage T4 recombination. The authors found a large number of branched molecules, joined at three-way and four-way DNA junctions. Alternatively, Holliday junctions have been detected in λ site-specific integration by the use of trapped recombination intermediates (Kitts & Nash, 1987; Nunes-Düby *et al.*, 1987; review by Leach, 1987). Other site-specific recombination systems with a similar dependence on localised homology, such as the phage P1 Cre-*lox* system (Hoess *et al.*, 1987) and the yeast FLP system (Jayaram *et al.*, 1988), also proceed via a Holliday junction intermediate; both of these studies utilised step-arrest mutations in recombinase enzymes.

It is possible that certain types of site-specific recombination occur by cleavage of cruciforms rather than Holliday junctions. Excision of a hybrid adeno-associated virus AAV/SV40 insert during an *in vitro* replication assay was found to be dependent on the presence of the palindromic AAV terminal repeat (Ward & Berns, 1991). The structure of the excision products is consistent with the resolution of a cruciform structure that is formed from the AAV terminal inverted repeats.

Holliday Junction Resolvases

The *E. coli* RuvC protein has been identified as a Holliday junction resolvase (Connolly *et al.*, 1991; Iwasaki *et al.*, 1991) that can cleave recombination intermediates which have been formed by RecA-mediated strand exchange (Dunderdale *et al.*, 1991). The *ruvC* gene maps as part of a locus that includes the SOS-inducible operon composed of the *ruvA* and *ruvB* genes. Together, the RuvA and RuvB proteins can interact with synthetic Holliday junctions to promote ATP-dependent branch migration (Parsons *et al.*, 1992; Tsaneva *et al.*, 1992), and exhibit DNA helicase activity *in vitro* (Tsaneva *et al.*, 1993). The combined action of RuvA and RuvB on a supercoiled plasmid leads to the reabsorption of an extruded cruciform (Shiba *et al.*, 1991). The biological roles of the RuvA, RuvB, and RuvC proteins are reviewed by West & Connolly (1992). The RecG protein of *E. coli* has recently been shown to bind specifically to synthetic Holliday junctions and, in the presence of ATP and Mg^{2+}, to dissociate these junctions (Lloyd & Sharples, 1993a). The enzymatic activities of the RecG and RuvAB proteins appear to overlap considerably, although RecG protein has a much higher specific activity for four-way DNA junctions (Lloyd & Sharples, 1993b). The movement and resolution of Holliday junctions by enzymes from *E. coli* are reviewed by Taylor (1992).

For a long time it was believed that the RecBCD enzyme of *E. coli* may constitute a Holliday junction resolvase. However, its has been shown that while RecBCD may cleave DNA strands at the base of cruciform structures, it must enter the termini of duplex DNA and approach the four-way junction from more than one direction in order to resolve it into recombinant products (Taylor & Smith, 1990). The authors propose that RecBCD may only cleave Holliday junctions in whose formation it participates. Nevertheless, a limited interaction of RecBCD with extruded cruciforms *in vivo* may be possible: methyl-directed mismatch repair of heteroduplex DNA that is capable of cruciform formation stimulates RecBCD-mediated recombination of an adjacent gene in *E. coli* (Kraczkiewicz-Dowjat & Fishel, 1990). The RecBCD enzyme of *E. coli* is reviewed by Taylor (1988).

The branched DNA structures observed by Broker & Lehman (1971) during phage T4 infection of *E. coli* are cleaved by the T4 endonuclease VII, the product of T4 gene *49*, which can resolve Holliday junctions (Mizuuchi *et al.*, 1982a). This enzyme specifically cleaves the crossover strands of synthetic Holliday junctions (Mueller *et al.*, 1988), and its activity is directed by local nucleotide sequence (Picksley *et al.*, 1990; Pottmeyer & Kemper, 1992). Phage T7 has a similar mode of

replication to phage T4, involving complex concatemeric intermediates, and also specifies a debranching enzyme (de Massy *et al.*, 1985). T7 endonuclease I, the product of the T7 gene *3*, has similarly been shown to cleave Holliday junctions (de Massy *et al.*, 1987) with a limited degree of site specificity (Picksley *et al.*, 1990).

Site-specific integration events that proceed via a four-way DNA junction intermediate are associated with resolving enzymes. However, in contrast to the other resolvases discussed here, these enzymes display a high degree of sequence specificity. Int, the phage-encoded integrase of λ, will resolve only four-way junctions that contain the *att* site (Hsu & Landy, 1984). Furthermore, Int protein shares many of the properties of type I topoisomerases. The Cre recombinase of bacteriophage P1, which is specific for *lox* sites (Hoess *et al.*, 1987), and the FLP protein of the yeast 2-μm plasmid (Meyer-Leon *et al.*, 1988; Jayaram *et al.*, 1988) are similar site-specific resolving enzymes.

Three different endonucleases capable of cleaving four-way DNA junctions have been purified from *S. cerevisiae*. A cruciform-specific activity was purified by West & Körner (1985), and its ability to resolve synthetic Holliday junctions has been shown to be dependent upon homologous DNA sequences (Parsons & West, 1988); branched DNA structures, which constitute substrates for T4 endonuclease VII, are not cut by the yeast enzyme. A second nuclease purified from *S. cerevisiae* (Symington & Kolodner, 1985) is also specific for four-way junctions as opposed to branched DNA (Evans & Kolodner, 1988). Encoded by the *CCE1* gene (Kleff *et al.*, 1992), it has been localised to the mitochondrial inner membrane (Ezekiel & Zassenhaus, 1993). Conversely, a third cruciform-specific enzyme from *S. cerevisiae* shows a similar activity to T4 endonuclease VII: cruciforms, four-way junctions, three-way junctions and heteroduplex loops are all cleaved (Jensch *et al.*, 1989).

Higher eukaryotes have been less forthcoming in the identification of Holliday junction resolvases. While cruciform-binding proteins have been identified in human extracts, these often show no endonucleolytic activity (Elborough & West, 1988). The resolution of synthetic Holliday structures by nucleases purified from human placental tissue (Jeyaseelan & Shanmugam, 1988) and HeLa cell extracts (Waldman & Liskay, 1988) has been demonstrated, but in neither case was it shown that cleavage occurred specifically at the four-way junction. An endonuclease activity from calf thymus has been shown to resolve synthetic Holliday junctions by specific cleavage in strands at opposite sides of the junction (Elborough & West, 1990). This mode of resolution is analogous to that catalysed by T4 endonuclease VII.

As described earlier, the structure of the four-way DNA junction is of great relevance to its cleavage by resolving enzymes. T4 endonuclease VII, the yeast endonuclease purified by West & Körner (1985), and the calf thymus enzyme purified by Elborough & West (1990) all interact with four-way DNA junctions on the minor groove side (upper side of Fig. 1.11). However, while T4 endonuclease VII cleaves the exchanging strands (dark grey in Fig. 1.11), the yeast and calf thymus enzymes cut the continuous strands (black in Fig. 1.11), 5' and 3' respectively to the point of strand exchange. The structure of DNA junctions and their interactions with resolvase enzymes are discussed by Duckett *et al.* (1992).

Cruciform DNA-Binding by HMG Boxes

The discovery of Holliday junction resolving enzymes has been paralleled by the characterisation of another eukaryotic protein that interacts with cruciform DNA. The specific binding of a protein from rat liver nuclei to four-way DNA junctions was reported by Bianchi (1988); this was later identified as the non-histone high mobility group protein 1 (HMG1) (Bianchi *et al.*, 1989), an abundant eukaryotic nuclear protein whose function is unclear. In contrast to the junction-specific enzymes described above, HMG1 has no nucleolytic activity (Bianchi, 1988). Furthermore, HMG1 protects cruciform structures from S1 nuclease digestion, and can remove the transcriptional block caused by cruciform extrusion in supercoiled DNA (Waga *et al.*, 1990). A protein with properties similar to HMG1 has recently been purified from *Ustilago maydis* (Kotani *et al.*, 1993) that is capable of enhancing DNA relaxation mediated by *U. maydis* type 1 topoisomerase (Thiyagarajan *et al.*, 1993).

The DNA binding site of HMG1 is composed of two highly similar segments, which have been termed HMG boxes (Bianchi *et al.*, 1992). Each of these segments is capable of specifically binding four-way DNA junctions, and exists as a dimer in solution. HMG1-like proteins in yeast and protozoa, and several eukaryotic transcription factors share the sequence motif defined by the HMG box. One of these transcription factors is SRY, the protein that determines the expression of male-specific genes in humans. Like HMG1, SRY recognises four-way DNA junctions irrespective of their sequence and additionally induces a sharp bend (~83°) in linear duplex DNA upon binding to its target site (Ferrari *et al.*, 1992), confirming that the interaction between HMG boxes and DNA is predominantly structure-specific. This conclusion has been reinforced by 2D ^{1}H-NMR spectroscopy of the HMG box region of rat HMG1: its structure is L-shaped and the angle of ~80° between the two arms is defined by a cluster of conserved residues (Weir *et al.*, 1993). This distinctive shape

may be significant in relation to the recognition of cruciforms; it is tempting to speculate that the L-shape of the HMG box may interact with the faces of the four-way DNA junction where acute angles of ~60° are exposed. The DNA-binding properties of HMG boxes are reviewed by Lilley (1992b).

A recent attempt to purify the main four-way DNA junction-binding protein of *E. coli* has identified the well-known histone-like protein HU as the prokaryotic analogue of eukaryotic HMG1 (Pontiggia *et al.*, 1993). The specific binding of protein HU to kinked DNA, a property it shares with HMG1, suggests a manipulation of DNA structure, in particular the axial distortion of the double helix. This would in turn facilitate other elements of DNA metabolism such as transcription, replication, recombination or repair. Histone H1, a major eukaryotic chromatin protein which may have a role as a repressor of transcription, has recently been shown to preferentially bind four-way DNA junctions (Varga-Weisz *et al.*, 1993).

DNA Replication

The development of anti-cruciform DNA monoclonal antibodies (Frappier *et al.*, 1987; Frappier *et al.*, 1989) has enabled the detection of cruciform structures *in situ*, and has provided evidence for their association with DNA replication in eukaryotic cells. The use of anti-cruciform DNA antibodies in a permeabilised cell system results in a 2 to 11-fold enhancement of DNA replication (Zannis-Hadjopoulos *et al.*, 1988); this was interpreted to be a consequence of antibody stabilisation of cruciforms located at or near replication origins. Fluorescent labelling of nuclei with these antibodies produces a non-uniform pattern of fluorescence in cells arrested at the G_1/S boundary, which changes as the cells are released from synchrony and pass through early S-phase (Ward *et al.*, 1990). Cruciforms could not be detected in metaphase chromosomes (Ward *et al.*, 1991). The relationship of the cruciform distribution with the nuclear matrix was examined by DNase I digestion (Ward *et al.*, 1991): the majority of cruciforms detectable at the G_1/S boundary and throughout the nucleus are readily digested by DNase I, whereas the nuclease-resistant fraction appears associated with the nuclear membrane and nucleolus. This is consistent with the idea that cruciforms *in vivo* are transient structures, which depend upon interactions with proteins essential for DNA replication. Anti-cruciform DNA antibodies have recently been used in the affinity purification of active mammalian origins of replication (Bell *et al.*, 1991).

Support for association of cruciform structures with origins of replication has been provided by DNA sequence analysis, in particular that of viral replication origins. As discussed in the section concerning palindromic DNA, bacteriophage λ has a near-perfect palindrome of 28 bp on the right of the origin that can assume a cruciform conformation (Hobom *et al.*, 1979). Similarly, the origin of SV40 houses a 27 bp perfect palindrome that includes the second T-antigen binding site (Tijan, 1978), while ori_{L2}, a replication origin of herpes simplex virus type 2, contains a 136 bp near-perfect palindrome (Lockshon & Galloway, 1986). Single-stranded regions have been detected in copies of *oriP*, the Epstein-Barr virus latent origin of plasmid replication, that are present within supercoiled plasmids (Orlowski & Miller, 1991). *oriP* contains 20 copies of a 30 bp near-palindromic repeat, which subsume a 12 bp palindromic consensus binding site for Epstein-Barr virus nuclear antigen (EBNA-1); the single-stranded structures are localised predominantly within this region. However, it is questionable whether EBNA-1 may interact with a target site in the cruciform conformation: formation of a hairpin structure by the 27 bp palindrome at the SV40 replication origin abolishes large T-antigen binding *in vitro* (Tenen *et al.*, 1983).

Cruciform extrusion at plasmid replication origins has been demonstrated *in vivo*. The leading strand replication origin of the staphylococcal plasmid pT181 consists of two adjacent inverted repeat elements which are involved in origin recognition by the initiator protein RepC (Wang *et al.*, 1993). Cruciform structures at the pT181 origin have been detected *in vitro* by S1 nuclease digestion and *in vivo* by bromoacetaldehyde treatment (Noirot *et al.*, 1990). Cruciform formation *in vitro* is enhanced by the binding of RepC, and the frequency of extrusion *in vivo* is correlated with the efficiency of RepC utilisation. This suggests that cruciform extrusion at the origin is involved in the initiation of pT181 rolling-circle replication.

Transcription

The model suggested by Gierer (1966), in which cruciform structures in operator sequences are the target of regulatory proteins, has remained attractive in spite of the discovery that most palindromic sequences in promoters are recognised by dimeric proteins when in the linear duplex conformation. The appeal of Gierer's model lies in the supercoiling dependence of both cruciform extrusion and the expression of a large number of genes. For example, if the 18 bp inverted repeat situated between the –10 and –35 consensus sequences in the promoter of the *gyrA* gene of *Klebsiella pneumoniae* (Dimri & Das, 1990) were to extrude as a cruciform, then it would interfere with initiation of transcription. This would provide a

mechanism for the homeostatic regulation of *gyrA*, encoding the DNA gyrase A subunit, by means of supercoiling-driven cruciform extrusion.

Support for this hypothesis has come from studies of promoter disruption by cruciform formation. Horwitz and Loeb (1988) constructed a promoter containing a 50 bp inverted repeat spanning from –23 to +27, with respect to the transcription start site +1; each repeat unit contained a –10 consensus sequence. Transcription from this promoter by RNA polymerase *in vitro* was repressed as the cruciform was extruded by increasing negative DNA supercoiling. Conversely, transcription *in vivo* was only induced as supercoiling was relaxed by the inhibition of DNA gyrase. A similar promoter, constructed by forming an inverted repeat from the –35 consensus sequence, also shows a supercoiling-dependent inhibition of transcription *in vitro* (Horwitz, 1989). Cruciform extrusion at this promoter additionally influences the expression of a distal gene, presumably by reducing the DNA superhelicity below the optimum needed for expression from the latter promoter. However, unlike the –10 inverted repeat, the –35 palindrome does not affect transcription *in vivo*.

While cruciform-mediated transcriptional blocks at the level of initiation *in vitro* have been reported elsewhere (Bagga *et al.*, 1990), the ability of cruciforms to impede transcription elongation has been questioned. Morales *et al.* (1990) found that cruciforms in supercoiled DNA present no hindrance to RNA polymerase *in vitro*, and are destabilised by movement of the transcription complex along the template. Furthermore, it has recently become clear that cruciform or hairpin formation may actually be required for promoter utilisation. Two sets of inverted repeats are present in each of the three promoters present in the coliphage N4 genome that are transcribed by the phage-encoded RNA polymerase. The upstream repeats are centred on a highly-conserved G/C-rich heptamer, and both specific sequences in these repeats and a hairpin structure in the template strand are required for promoter recognition (Glucksmann *et al.*, 1992). The hairpin formed by these inverted repeats appears to be stabilised by the binding of *E. coli* single-stranded DNA-binding (SSB) protein to yield an activated promoter. Similarly, cruciform extrusion by a 23 bp imperfect palindrome in the enhancer region of the human encephalin gene may play a role in the latter's cAMP-inducible activation (McMurray *et al.*, 1991).

Significant contributions to an understanding of the effects of DNA supercoiling on transcription, and vice versa, have come from studies of cruciform formation by $(A\text{-}T)_n$ sequences. These sequences have no detectable kinetic barrier to extrusion (Greaves *et al.*, 1985), and in this respect cruciform formation by $(A\text{-}T)_n$

sequences resembles Z-DNA formation by $(C\text{-}G)_n$ sequences, which has been widely used as a probe for localised DNA supercoiling *in vivo* (review by Rahmouni, 1992). Moreover, the spontaneous unwinding of promoter and terminator regions in negatively supercoiled DNA (Drew *et al.*, 1985; Sheflin & Kowalski, 1985) would help overcome the activation energy of cruciform extrusion by $(A\text{-}T)_n$ sequences. The *in situ* chemical probing of extrusion by $(A\text{-}T)_n$ sequences has revealed that the superhelical tension of cellular DNA responds directly to environmental factors (McClellan *et al.*, 1990; Dayn *et al.*, 1991). In particular, the effect of osmotic shock is an elevation of negative DNA supercoiling, which can be detected by cruciform formation. This increase in supercoiling affects the expression of a number of genes such as the *proU* locus, which encodes an osmotically inducible glycine betaine transport system that is important in the adaptation to osmotic stress (Higgins *et al.*, 1988; review by Higgins *et al.*, 1990a). It has recently been established that the histone-like protein H1 (H-NS) (reviewed by Higgins *et al.*, 1990b) plays an important role in determining the level of chromosomal supercoiling that underlies the osmotic-dependent regulation of gene expression (Hulton *et al.*, 1990).

The measurement of *in vivo* cruciform extrusion from $(A\text{-}T)_n$ sequences by *in situ* chemical probing has also been instrumental in determining the effect of transcription on DNA supercoiling. Cruciform formation occurs more often upstream than downstream of transcriptional units, and is particularly frequent where $(A\text{-}T)_n$ sequences are located between two divergent promoters (Zheng *et al.*, 1991; Dayn *et al.*, 1992). These results, indicating that transcriptionally-driven negative DNA supercoiling upstream of promoters may cause cruciform extrusion, conform to the twin supercoiling domain model of Liu & Wang (1987). DNA upstream of two divergent promoters will become negatively supercoiled by the movement of RNA polymerase along the transcript; resistance to the rotational motion of the transcription ensemble is large, and consequently the advancing polymerase generates positive supercoils ahead of it, and negative supercoils behind it (Wu *et al.*, 1988). The modulation of DNA supercoiling by transcription may catalyse cruciform extrusion from palindromic sequences located in promoter regions, which may in turn prevent initiation at that promoter.

Such a mode of regulation may exist in chicken mitochondrial DNA, where transcription initiates predominantly from one bidirectional promoter which contains a near-perfect (A+T)-rich palindrome of 18 bp (L'Abbé *et al.*, 1991). A similar arrangement is found at the bidirectional *mob* promoter of *Thiobacillus ferrooxidans* plasmid pTF1, where a 23 bp near-perfect palindrome located in the promoter is also

contained within the overlapping *oriT* origin of conjugative DNA transfer (Drolet & Lau, 1992). Binding of MobL, one of the gene products transcribed from the *mob* promoter, to the *oriT* region *in vitro* is dependent on the formation of a hairpin structure by the palindromic sequence. In this case, transcriptionally-driven negative supercoiling of the upstream region may promote cruciform extrusion and thereby facilitate the recognition of the transfer origin by MobL.

Genetic Disease

Some evidence, albeit conjectural, links the existence of cruciform structures to genetic disease. Cruciform DNA, by virtue of its unusual structure, may be a preferential target of mutation, and may additionally be refractory to DNA repair. The evolutionary consequences of non-random DNA damage and repair are discussed by Boulikas (1992). Alternatively, cruciform structures may themselves elicit mutagenic events: a molecular analysis of DNA junctions produced by illegitimate recombination in human cells has revealed some association with palindromes, (A+T)-rich regions, alternating purine/pyrimidine sequences and *Alu* family repeats (Stary & Sarasin, 1992). All of these sequences are capable of forming unusual DNA secondary structures. Furthermore, long inverted repeats have been associated with the structural instability of chromosome 15 in Prader-Willi syndrome (Donlon *et al.*, 1986); the chromosomal aberrations may be a consequence of cruciform extrusion. Recent analyses of tumour-associated genetic changes, such as DNA translocation or deletion, have implicated DNA repeats and associated secondary structures in genomic instability and neoplasia (Bouffler *et al.*, 1993). The involvement of DNA structure in mutation and genetic disease is reviewed by Sinden & Wells (1992).

Concluding Remarks

In spite of the numerous examples of cruciform formation described above, it should be remembered that the presence of a DNA hairpin does not constitute evidence for cruciform extrusion. This structural transition may be catalysed *in vivo* by DNA unwinding, generating single-stranded regions which are capable of forming a cruciform by simple closure. DNA and RNA polymerases may be responsible for the DNA unwinding, thereby overcoming the kinetic barrier to cruciform extrusion. This would account for the frequent association of cruciform structures with origins of replication and transcription promoters. The uncertainty arising from the question of whether an elevated level of negative supercoiling, or merely extensive DNA

unwinding is responsible for cruciform formation disqualifies much of the work described above from constituting bona fide examples of cruciform extrusion *in vivo*.

Extrusion by the S-type mechanism outlined earlier has never been clearly demonstrated *in vivo*. The accounts of cruciform extrusion *in vivo* that have been reported to date nearly all involve highly (A+T)-rich sequences that would not extrude in a centre-dependent manner (e.g. del Olmo & Pérez-Ortín, 1993). Even the most celebrated report of cruciform formation *in vivo* detected extrusion from a palindrome embedded in a region of 72% (A+T)-rich DNA (Panayotatos & Fontaine, 1987); under these circumstances, a C-type mechanism may operate. Evidence for cruciform extrusion from inverted repeats of a normal base composition, which are not flanked by (A+T)-rich inducing sequences, has generally been negative (e.g. Sinden *et al.*, 1983). It may be the case, as suggested by Courey & Wang (1983), that cruciform formation by an S-type mechanism is kinetically forbidden under physiological conditions. However, the work presented in this thesis provides evidence for cruciform extrusion *in vivo* by a centre-dependent pathway that is consistent with the S-type mechanism.

CONCLUSION

Palindromic DNA sequences are widespread throughout the genomes of all organisms examined for their presence. Their inverted symmetry makes them ideal target sites for regulatory proteins, which may interact with the palindrome as a linear or a hairpin structure. In spite of this universal role for palindromic DNA, its distribution is not uniform: long inverted repeats, while common in the genomes of eukaryotes, are completely absent in prokaryotic organisms. This disparity is intimately associated with the unusual behaviour of long inverted repeats in prokaryotes. Perfect palindromes of greater than ~100 – 200 bp are unstable in wild-type *E. coli*, frequently using direct repeats as deletion endpoints, and confer inviability upon their host vector. This latter property may be circumvented by the use of *sbcCD* hosts, implying an interaction of the SbcCD protein with such sequences.

The impetus for this unusual behaviour of long DNA palindromes in *E. coli* may be provided by cruciform extrusion. The formation of cruciform structures is thermodynamically favoured in negatively supercoiled DNA, although considerable kinetic barriers exist. Detailed analysis of the kinetics of cruciform extrusion *in vitro* has led to the proposal of two mechanisms: the C-type pathway, which is restricted to palindromes in the sequence context of (A+T)-rich DNA under conditions of low ionic strength, and the centre-dependent S-type pathway which operates at physiological salt concentrations for palindromes of a normal base composition. The investigation of cruciform extrusion *in vitro* has been made possible by a wide variety of techniques, only some of which can be applied *in vivo*. While the formation of cruciform structures *in vivo* by $(A\text{-}T)_n$ sequences, which have no kinetic barrier to extrusion, has been demonstrated conclusively, evidence of cruciform extrusion by palindromes of a normal base composition has been less forthcoming.

This thesis concerns the queries raised by the findings summarised above. Firstly, it attempts to answer the lingering question of cruciform extrusion *in vivo* from long DNA palindromes of an average base composition. The stable propagation of a 476 bp perfect palindrome was permitted by the use of a λ *red gam* χ^+ vector in conjunction with *E. coli sbcC* and *recBC sbcBC* hosts. Cruciform formation was measured *in vivo* by the inhibition of methylation by endogenous Dam methylase at a target site located at the centre of the palindrome. This non-invasive technique has been developed only recently, and elements of the method presented here may be applicable to other studies of DNA structural transitions, and to the investigation of protein-DNA interactions *in vivo*.

Secondly, this thesis explores the relevance of the S-type pathway to cruciform extrusion *in vivo*. A set of four palindromes was constructed, differing symmetrically at only one central position yet retaining the Dam methylase target site (GATC); the general sequence of the palindrome centre is given below.

```
5' ...AGCTGTNGATCnACAGCT... 3'
3' ...TCGACAnGATCNTGTCGA... 5'
```

These palindromes were then used in the assay for cruciform extrusion *in vivo* outlined above The frequency of cruciform formation was compared to the theoretically-determined stability of these central DNA sequences, which is predicted by the S-type mechanism to be the major rate-determinant of extrusion. The influence on cruciform extrusion *in vivo* of a small central interruption to the inverted symmetry was also evaluated.

Lastly, the biological consequences of cruciform extrusion *in vivo* are addressed. The viability of λ phage carrying the concerted set of palindromes was quantified in terms of their plaque size on a lawn of *E. coli sbcC* cells. Plaque area is a measure of the phage burst size, and consequently describes its viability in that host strain. The plaque sizes were compared to the frequencies of *in vivo* cruciform extrusion, and both sets of figures were evaluated in the light of current models for palindrome-mediated inviability.

It is hoped that the methods and results described in this thesis will benefit research in a number of fields. The demonstration of cruciform extrusion *in vivo* will contribute to investigations into DNA secondary structure transitions, and assist in the characterisation of palindrome-mediated DNA instability and vector inviability. A detailed knowledge of these phenomena may expedite the analysis of those intractable eukaryotic DNA sequences that have been refractory to cloning in *E. coli*. The palindrome may thus finally make a lasting contribution to the welfare of mankind that extends beyond illuminating phrases such as "Madam I'm Adam".

CHAPTER 2

MATERIALS & METHODS

SECTION 1

MATERIALS

Microbiological Strains, Media and Solutions

Strains

All bacterial strains, bacteriophage λ strains and plasmid vectors used in this work are described in Tables 2.1, 2.2 and 2.3 respectively.

Media

Bacteriological Media

The following quantities are for 1 litre final volumes made up in distilled water and sterilised by autoclaving for 20 minutes at 15 lb in^{-2}.

BBL Agar

10 g trypticase (Baltimore Biological Laboratories), 5 g NaCl, 10 g Bacto-agar (Difco), adjusted to pH 7.2 with NaOH.

BBL Top Agar

As BBL agar, but containing only 6.5 g Bacto-agar (Difco) per litre.

BBL Agar (Plaque Area)

10 g Select® trypticase (Becton Dickinson), 14 g NaCl, 10 g Bacto-agar (Difco), 10 mM Tris/HCl (pH 7.5).

BBL Top Agar (Plaque Area)

As BBL agar (plaque area), but containing only 5 g Bacto-agar (Difco) per litre.

L Agar

10 g Bacto-tryptone (Difco), 5 g yeast extract (Difco), 10 g NaCl, 15 g Bacto-agar (Difco), adjusted to pH 7.2 with NaOH.

L Broth

10 g Bacto-tryptone (Difco), 5 g yeast extract (Difco), 10 g NaCl, adjusted to pH 7.2 with NaOH.

L Broth p/c

L broth supplemented with 0.2% maltose and 5 mM $MgSO_4$.

L Broth (Liquid Lysis)

10 g Bacto-tryptone (Difco), 5 g yeast extract (Difco), 10 g NaCl, 10 mM $MgSO_4$, 20 mM Tris/HCl (pH 7.5).

Phage Buffer

3 g KH_2PO_4, 7 g Na_2HPO_4, 5 g NaCl, 1 mM $MgSO_4$, 1 mM $CaCl_2$, 1 ml gelatin (1% w/v).

Seakem Agar

L broth containing 12 g Seakem agarose (FMC Bioproducts) per litre, and supplemented with 0.3% glucose, 2 mM $MgSO_4$, 0.8 mM $CaCl_2$, 4 µM $FeCl_3$, 10 mM Tris/HCl (pH 7.5), 10 mg vitamin B_1, after cooling to 46°C. Seakem agarose is a pure form of agarose containing no contaminating enzyme inhibitors.

Seakem Top Agar

L broth containing 4 g Seakem agarose (FMC Bioproducts) per litre.

TM Buffer

10 mM Tris/HCl (pH 7.5), 10 mM $MgSO_4$.

Media Additives

0.5 M $CaCl_2$ Stock

Made up in distilled water, autoclaved.

0.01 M $FeCl_3$ Stock

Made up in distilled water, filter sterilised.

1 M $MgSO_4$ Stock

Made up in distilled water, autoclaved.

20% Glucose Stock

Made up in distilled water, filter sterilised.

20% Maltose Stock

Made up in distilled water, filter sterilised.

Ampicillin (100 mg ml^{-1})

Ampicillin (Beecham Pharmaceuticals) was stored at –20°C and used at 100 µg ml^{-1}.

Chloramphenicol (20 mg ml^{-1})

Chloramphenicol (Sigma Chemical Company) was made up in 100% ethanol and stored at –20°C. It was used at 50 µg ml^{-1}.

Tetracycline (15 mg ml^{-1})

Tetracycline (Sigma Chemical Company) was made up in 50% (v/v) ethanol and stored at –20°C. It was used at 10 µg ml^{-1}.

Xgal (25 mg ml^{-1})

Xgal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) (Northumbria Biologicals) was made up in dimethylformamide and stored at –20°C. It was used at 40 μg ml^{-1}.

Vitamin B_1 (5 mg ml^{-1})

Vitamin B_1 (Sigma Chemical Company) was made up in distilled water, filter sterilised and stored at 4°C.

Solutions

Solutions for Purification of Bacteriophage λ Particles

DNase I (10 mg ml^{-1})

DNase I (Sigma Chemical Company) was made up in sterile 50 mM Tris/HCl (pH 7.5), 10 mM NaCl, 1 mM DTT, 100 μg ml^{-1} BSA, 50% (w/v) glycerol and stored at –20°C.

RNase A (10 mg ml^{-1})

RNase A (Sigma Chemical Company) was made up in sterile 10 mM Tris/HCl (pH 7.5), 15 mM NaCl, then placed in a boiling water bath for 10 minutes and allowed to cool slowly to room temperature. This treatment denatures contaminating DNases but not RNase A. The RNase A was stored at –20°C.

30% PEG 6000, 3 M NaCl

PEG 6000 (Fluka Chemie) was made up in sterile 3 M NaCl and stored at 4°C. High concentrations of PEG 6000 cannot be autoclaved or filter sterilised.

CsCl Solutions

CsCl (Boehringer Mannheim) was made up in sterile phage buffer to concentrations of 31.2%, 45.4% and 56.2% (w/w). The CsCl causes gelatin in the phage buffer to precipitate. The solutions were therefore left overnight for the precipitate to settle.

Solutions for Transformation of *Escherichia coli*

0.1 M $CaCl_2$

Made up in distilled water, autoclaved.

Solutions for *in Vitro* Packaging of λ DNA

These solutions were not made up in the course of this work, as ample pre-existing stocks were available (stored at –70°C). The composition of these solutions are given

only for reference; the stock solutions listed are not described elsewhere. The methods of preparation are similarly abbreviated.

Buffer A

20 mM Tris/HCl (pH 7.5), 3 mM $MgCl_2$, 0.05% (v/v) β-mercaptoethanol, 1 mM EDTA, stored at –70°C.

Buffer M1

6 mM Tris/HCl (pH 7.5), 30 mM spermidine, 60 mM putrescine, 18 mM $MgCl_2$, 15 mM ATP, 0.2% (v/v) β-mercaptoethanol, stored at –70°C.

Sonicated Extract

Lysogenic *E. coli* BHB2690 was grown shaking in 500 ml L broth at 30°C to OD_{650} = 0.3. Lysogens were induced by transferring the culture to a 43°C shaking waterbath for 15 minutes, followed by vigorous shaking at 37°C for 1 hour. After cooling on ice/water for 10 minutes, the cell suspension was centrifuged at 6 krpm for 6 minutes at 4°C (Sorvall centrifuge, GSA rotor). Cell pellets were resuspended in 1/500 volume buffer A, transferred to a single 30 ml Nalgene polypropylene centrifuge tube and diluted with 2.6 ml buffer A. The suspension was sonicated in ~3 second bursts (to prevent foaming) until no longer viscous and centrifuged at 6 krpm for 6 minutes at 4°C (Sorvall centrifuge, SS34 rotor). 50 µl of supernatant was aliquoted to pre-cooled Eppendorf tubes, frozen on liquid N_2, and stored at –70°C.

Freeze-Thaw Lysate

Lysogenic *E. coli* BHB2688 was grown shaking in 500 ml L broth at 30°C to OD_{650} = 0.3. Lysogens were induced by transferring the culture to a 43°C shaking waterbath for 15 minutes, followed by vigorous shaking at 37°C for 1 hour. After cooling on ice/water for 10 minutes, the cell suspension was centrifuged at 6 krpm for 6 minutes at 4°C (Sorvall centrifuge, GSA rotor). Cell pellets were resuspended in 1/100 volume cold 10% sucrose, 50 mM Tris/HCl (pH 7.5), pooled and dispensed into a Nalgene ultracentrifuge tube. 150 µl freshly prepared lysozyme (10 mg ml^{-1} in 0.25 M Tris/HCl, pH 7.5) was added and gently mixed. The suspension was frozen on liquid N_2, thawed first at room temperature and then at 4°C until the pellet is completely thawed. The suspension was kept on ice, 150 µl of buffer M1 was added and gently mixed, and then centrifuged at 40 krpm for 1 hour at 4°C (Sorvall ultracentrifuge, Ti50 rotor). 55 µl of supernatant was aliquoted to pre-cooled Eppendorf tubes, frozen on liquid N_2, and stored at –70°C.

Table 2.1

***Escherichia coli* Strains Used in this Work.**

1. The plasmids in these strains encode *Eco*RI restriction/modification (Table 2.3).
2. This strain is a derivative of K802, F⁻ *lacY*1 *supE*44 *galK*2 *galT*22 *mcrA* *rfbD*1 *metB*1 *mcrB*1 *hsdR*2 (Wood, 1966).
3. This strain is a derivative of JM83, *ara* Δ(*lac-proAB*) *rpsL* (Str^r) ϕ80 *lacZ*ΔM15 (Yanisch-Perron *et al.*, 1985).
4. These strains are derivatives of AB1157, F⁻ *thi*-1 *his*-4 Δ(*gpt-proA*)62 *argE*3 *thr*-1 *leuB*6 *kdgK*51 *rfbD*1 (?) *ara*-14 *lacY*1 *galK*2 *xyl*-5 *mtl*-1 *tsx*-33 *supE*44 *rpsL*31 (Howard-Flanders & Theriot, 1966).
5. These strains are recombinants from a mating between a *pro*$^+$ *lac*$^+$ derivative of KL425 (Lloyd *et al.*, 1987) and N2373, Hfr (Cavalli, PO2A) *sbcC*201 (Lloyd, 1991). The strains were subsequently made *thy*$^+$ by P1 transduction from AC301 (Lloyd *et al.*, 1988).
6. These strains are *pro*$^+$ recombinants from a mating between N1116 and a *pro*$^+$ *sbcC*201 transductant of χ342 (CGSC4515, Hfr (Cavalli, PO2A) *proC*29 *metB*1 *relA*1) (Lloyd & Buckman, 1985). The *recA*269::Tn*10* mutation was subsequently introduced by P1 transduction.
7. This strain is a derivative of W3110, F⁻ IN(*rrnD-rrnE*) *thyA*36 (Bachmann, 1972).
8. The plasmid in this strain carries the λ *gam* gene (Table 2.3).

Table 2.1

Strain	Description	Reference/Source	Notes
594	F⁻ *lac galK2 galT22 rpsL179* (Str^r)	Campbell (1965)	
DL558	HB101 / NTP14 pLV56	D. Leach	1
DL559	HB101 / pLV59	D. Leach	1
DL655	*recA*::Kan^r *sbcC201*	D. Leach	2
DL733	Δ*sbcCD*::Kan^r	D. Leach	3
HB101	F⁻ Δ(*mcrC-mrr*) *leu supE44 ara14 galK2 lacY1 proA2 rpsL20* (Str^r) *xyl-5 mtl-1 recA13*	Boyer & Roulland-Dussoix (1969)	
JC7623	*recB21 recC22 sbcB15 sbcC201*	Kusher *et al.* (1971)	4
JC9387	*recB21 recC22 sbcB15 sbcC201 sup°*	F. Stahl	4
JL32	*recB21 recC22 sbcB15 sbcC201 sup° dam3* (Str^s) *xyl*$^+$	J. Lindsey (Ph.D. Thesis, 1987)	4
LR299	*ruvC53 sbcC201 phoR79*::Tn*10 eda* (?)	R. Lloyd	4
LR327	*ruvC53 sbcC201 recG258*::Kan^r	R. Lloyd	4
N2361	As AB1157 (see Note 4)	Lloyd & Buckman (1985)	4
N2364	*sbcC201 phoR79*::Tn*10*	Lloyd & Buckman (1985)	4
N2671	F⁻ *thi-1 metE70 leuB6 cysC43 ara-14 mtl-1 xyl-5 str-109 spc-15 nalA*⁻ *supD*⁻ *thy*$^+$ *recD*$^+$	R. Lloyd	5
N2673	As N2671, except *sbcC201*	R. Lloyd	5
N2691	F⁻ *his-4 argE3 leuB6 ara-14 lacY1 galK2 xyl-5* (Str^r) *mtl-1 supE44 recA269*::Tn*10*	Chalker *et al.* (1988)	4, 6
N2693	As N2691, except *lac*$^+$ *sbcC201*	Chalker *et al.* (1988)	4, 6
NM701	*mcrA* Δ(*mrr-hsdRMS-mcrBC*) *sup° recA56* / p*gam*	N. Murray	7, 8

Table 2.2

Bacteriophage λ Strains Used in this Work.

1. *spi*6 and *b*1453 are deletions of λ which render the phage *red gam*.
2. *cI*857 encodes a temperature-sensitive repressor, while *cIam* is a suppressible mutation in the *cI* repressor gene.
3. Δ*B* is a deletion of the *Eco*RI B fragment of λ.
4. λDRL116 was not used in this work but is described in detail, and is therefore included in Table 2.2.
5. *pal*571 indicates that the phage carries a 571 bp asymmetric palindrome (Fig. 4.2).
6. *pal*462(*Sac*I) indicates that the phage carries a 462 bp perfect palindrome with a central *Sac*I site (see Fig 4.2 and Appendix).
7. e.g. *pal*476(*Bam*HI) indicates that the phage carries a 476 bp perfect palindrome with a central *Bam*HI site, created by an oligonucleotide insertion at the centre of the *pal*462(*Sac*I) palindrome (see Chapters 4, 8 and 9).
8. λ χ^+*C*153 phage have a χ^+ site at bp 38481 – 38488 enabling efficient replication of *red gam* phage in a *recBCD*$^+$ host (Smith, 1988b) (see Chapter 5).
9. λDRL167 was constructed by a cross of λDRL133 and λDRL152.
10. λDRL186 was constructed by a cross of λDRL151 and λNEM1239.
11. λDRL186 and λNEM1239 appear to have a χ^+ site, but its location is uncertain (see Chapter 5).
12. *b*538 is a deletion of bp 20809 – 29093 of λ and removes *Eco*RI sites 1 and 2.
13. e.g. *srI*3° is a mutation destroying *Eco*RI site 3 of λ.
14. *Nam*7 and *Nam*53 are suppressible mutations in *N*, the gene encoding an antitermination factor required for transcription from p_L and p_R (see Chapter 7).
15. *nin*5 is a deletion of bp 40502 – 43307 of λ which removes the t_{R2} terminator, enabling N-independent growth (see Chapter 7).

Table 2.2

Strain	Description	Source	Notes
λDRL112	*spi*6 *cI*857	D. Leach	1, 2, 3
λDRL116	Δ*B pal*571 *spi*6 *cI*857	D. Leach	1, 2, 3, 4, 5
λDRL133	Δ*B pal*462(*Sac*I) *spi*6 *cI*857	D. Leach	1, 2, 3, 6
λDRL151	Δ*B pal*476(*Bam*HI) *spi*6 *cI*857	This work	1, 2, 3, 7
λDRL152	*spi*6 *cI*857 χ^+C153	D. Leach	1, 2, 8
λDRL167	Δ*B pal*462(*Sac*I) *spi*6 *cI*857 χ^+C153	A. Davison	1, 2, 3, 6, 8, 9
λDRL176	Δ*B pal*476(*Bam*HI) *spi*6 *cI*857 χ^+C153	This work	1, 2, 3, 7, 8
λDRL177	Δ*B pal*476(*Bcl*I) *spi*6 *cI*857 χ^+C153	This work	1, 2, 3, 7, 8
λDRL178	Δ*B pal*476(*Bgl*II) *spi*6 *cI*857 χ^+C153	This work	1, 2, 3, 7, 8
λDRL179	Δ*B pal*476(*Pvu*I) *spi*6 *cI*857 χ^+C153	This work	1, 2, 3, 7, 8
λDRL180	Δ*B pal*476(*Xho*II) *spi*6 *cI*857 χ^+C153	This work	1, 2, 3, 7, 8
λDRL186	Δ*B pal*476(*Bam*HI) *spi*6 *cI*857 χ^+(?)	This work	1, 2, 3, 7, 10, 11
λMMS659	*b*1453 *cI*857	F. Stahl	1, 2
λNEM43	*b*538 *cIam*	N. Murray	2, 12
λNEM378	*b*538 *cIam srI*3° *srI*4° *srI*5°	N. Murray	2, 12, 13
λNEM423	*Nam*7 *Nam*53 *cI*857 *nin*5	N. Murray	2, 14, 15
λNEM1239	*b*1453 *cI*857 *srI*4° *nin*5 *srI*5° χ^+(?)	N. Murray	1, 2, 11, 13, 15

Table 2.3

Plasmid Vectors Used in this Work.

1. Encodes wild-type *Eco*RI restriction and modification enzymes (see Chapter 3).
2. Derivatives of pMS2B containing 28 bp *Taq*I (or *Bst*BI) palindrome centre fragment inserted at the *Acc*I site (see Fig. 11.1).
3. Expresses λ *gam* gene under control of the λ late promoter $p_{R'}$ (see Chapter 7).
4. Encodes wild-type *Eco*RI methylase (see Chapter 3).
5. Encodes temperature-sensitive *Eco*RI methylase and inactive (?) *Eco*RI restriction endonuclease (see Chapter 3).
6. Derivative of pUC18 (Norrander *et al.*, 1983; Yanisch-Perron *et al.*, 1985) which for unknown reasons stabilises long DNA palindromes (C. Blake, M. Shaw & D. Leach, unpublished observations) (see Chapter 11).
7. Derivative of pBR322 containing tandem insertions of a $(CG)_{13}GATC(CG)_{13}$ sequence at the *Eco*RI site (see Chapter 7).

Table 2.3

Plasmid	Markers	Reference/Source	Notes
NTP14	Amp^r	Smith *et al.* (1976)	1
pBR322	Amp^r Tet^r	Bolivar *et al.* (1977b)	
pDRL151	Amp^r	This work	2
pDRL176	Amp^r	This work	2
pDRL177	Amp^r	This work	2
pDRL178	Amp^r	This work	2
pDRL179	Amp^r	This work	2
pDRL180	Amp^r	This work	2
p*gam*	Amp^r	Crouse (1985)	3
pLV56	Chl^r	O'Connor & Humphreys (1982)	4
pLV59	Chl^r	O'Connor & Humphreys (1982)	5
pMS2B	Amp^r	M. Shaw	6
pZTA1	Amp^r Tet^r	This work	7

Materials for DNA Purification and Manipulation

General Solutions and Materials for DNA Purification

Unless otherwise stated, general laboratory chemicals were purchased from Sigma Chemical Company, Fisons or BDH.

1 M Tris/HCl Stock (pH 7.5)

1 M Tris base, adjusted to pH 7.5 with concentrated HCl, autoclaved.

0.5 M EDTA Stock (pH 8)

0.5 M EDTA disodium salt, adjusted to pH 8 with glacial acetic acid, autoclaved.

TE Buffer Stock (10×)

100 mM Tris, 10 mM EDTA, adjusted to pH 7.5 with concentrated HCl, autoclaved.

Solutions for Phenol, Phenol-Chloroform and Chloroform Extraction

Phenol

Distilled, liquefied (88%) phenol (Rathburn Chemicals) was stored in 25 ml aliquots at –20°C, and protected from light in foil-wrapped 25 ml polypropylene tubes. After thawing at room temperature, 0.1% (w/v) 8-hydroxyquinoline (Sigma Chemical Company) was added and the phenol equilibrated using TE buffer. Initially, this was performed by emulsifying the phenol with 10× TE buffer (100 mM Tris/HCl, pH 7.5, 10 mM EDTA) for 5 minutes, followed by centrifugation at 4.5 krpm in an MSE Centaur-2 bench centrifuge. The aqueous layer was removed, and equilibration repeated another two times using 1× TE buffer. The aqueous layer was removed and replaced with 2 ml 1× TE buffer 0.2% (v/v) β-mercaptoethanol (Sigma Chemical Company). The phenol was protected from light, stored at 4°C in a 25 ml polypropylene tube, and used within one month.

Phenol-Chloroform

As above, except after final equilibration the phenolic phase was combined with an equal volume of 24:1 chloroform-isoamylalcohol (see below) and centrifuged at 4.5 krpm in an MSE Centaur-2 bench centrifuge. The aqueous layer was removed and replaced with 2 ml 1× TE buffer 0.2% (v/v) β-mercaptoethanol as above. The phenol-chloroform was protected from light, stored at 4°C and used within one month.

24:1 Chloroform-isoamylalcohol

24 volumes of chloroform were combined with 1 volume of isoamylalcohol and protected from light in a brown glass bottle.

Solutions for Ethanol and Isopropanol Precipitation

3 M Sodium Acetate (pH 5.3)

Sodium acetate solution (3 M in acetate) was prepared by adding 0.19 volumes of sterile 3 M acetic acid to 0.81 volumes of sterile 3 M sodium acetate, autoclaved.

3 M Sodium Acetate (pH 7.0)

3 M sodium acetate, adjusted to pH 7.0 with glacial acetic acid, autoclaved.

70% Ethanol

Prepared by adding 0.3 volumes of sterile distilled water to 0.7 volumes of ethanol.

Preparation of Dialysis Tubing

Dialysis tubing was cut into lengths of 20 cm and boiled for 10 minutes in 2 *l* of 2% sodium bicarbonate and 1 mM EDTA. It was then rinsed thoroughly in distilled water and boiled for 10 minutes in distilled water. After cooling, the tubing was stored at 4°C submerged in 1 mM EDTA 50% (v/v) ethanol. Before use, the tubing was washed inside and out with sterile distilled water.

Materials for Purification of DNA from Agarose Gels and Solutions

Purification of DNA from agarose gels and solutions was performed using GENECLEAN® and MERmaid™ kits (BIO 101). The solutions used were those provided by the manufacturer, and are listed here only for reference.

6 M NaI Stock

6 M NaI, protected from light at 4°C.

GLASSMILK®

Suspension of silica matrix in distilled water, stored at 4°C.

NEW Wash

20 mM Tris/HCl (pH 7.2), 0.2 M NaCl, 2 mM EDTA, 50% ethanol, stored at –20°C.

High Salt Binding Solution

Saturated solution of sodium perchlorate, stored at room temperature.

GLASSFOG™

Suspension of fine silica-based matrix in distilled water, stored at room temperature.

Ethanol Wash

20 mM Tris/HCl (pH 7.2), 0.2 M NaCl, 2 mM EDTA, 90% ethanol, stored at room temperature.

Solutions for Bacteriophage λ DNA Purification

Small Scale Method

Solutions for the small scale λ DNA purification method are all listed elsewhere.

Large Scale Method

Pronase (20 mg ml^{-1})

Pronase (serine protease from *Streptomyces griseus*) (Sigma Chemical Company) was prepared in 10 mM Tris/HCl (pH 7.5), 10 mM NaCl and autodigested for 1 hour at 37°C. Stored at –20°C.

Pronase Dialysis Buffer

20 mM Tris/HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, 0.002% Triton X-100.

Solutions for Plasmid DNA Purification

Small Scale Method

TEG

0.9% glucose, 25 mM Tris/HCl (pH 7.5), 10 mM EDTA, filter sterilised.

Alkaline SDS

0.2 M NaOH, 1% SDS (sodium dodecyl sulphate), freshly prepared.

Potassium Acetate

Potassium acetate (3 M in potassium, 5 M in acetate) was prepared by adding 0.4 volumes of sterile 5 M acetic acid to 0.6 volumes of sterile 5 M potassium acetate, autoclaved.

TE-RNase A

10 mM Tris/HCl (pH 7.5), 1 mM EDTA, 20 μg ml^{-1} RNase A. Prepared by adding 100 μl RNase A stock (10 mg ml^{-1}) to 50 ml sterile 1× TE buffer. Stored at 4°C.

Large Scale Method

Large scale plasmid DNA preparation was performed using a QIAGEN plasmid midi kit. The solutions used were those provided by the manufacturer, and are listed here only for reference.

Buffer P1

100 μg ml^{-1} RNase A, 50 mM Tris/HCl (pH 8.0), 10 mM EDTA, stored at 4°C.

Buffer P2

200 mM NaOH, 1% SDS.

Buffer P3

3 M potassium acetate (pH 5.5), stored at 4°C.

Buffer QBT

750 mM NaCl, 50 mM MOPS (pH 7.0), 15% ethanol, 0.15% Triton X-100.

Buffer QC

1 M NaCl, 50 mM MOPS (pH 7.0), 15% ethanol.

Buffer QF

1.25 M NaCl, 50 mM Tris/HCl (pH 8.5), 15% ethanol.

Enzymes and Buffers for DNA Manipulation

Restriction Endonucleases

All restriction endonucleases used in this work and their incubation buffers are described in Tables 2.4 and 2.5 respectively.

DNA Sequencing

DNA sequencing was performed using a Sequenase® v2.0 DNA sequencing kit (United States Biochemical). With the exception of the [α^{35}S]dATP (Amersham) and the sequencing primer, the solutions used were those provided by the manufacturer, and are listed here only for reference. All solutions were stored at –20°C.

Sequenase® Buffer (5×)

200 mM Tris/HCl (pH 7.5), 100 mM $MgCl_2$, 250 mM NaCl.

DTT

0.1 M DTT (dithiothreitol), prepared in distilled water.

Labelling Mix (5×)

7.5 µM dGTP, 7.5 µM dCTP, 7.5 µM dTTP.

ddG Termination Mix

80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddGTP, 50 mM NaCl

ddA Termination Mix

80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddATP, 50 mM NaCl

ddT Termination Mix

80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddTTP, 50 mM NaCl

ddC Termination Mix

80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddCTP, 50 mM NaCl

Enzyme Dilution Buffer

10 mM Tris/HCl (pH 7.5), 5 mM DTT, 0.5 mg ml^{-1} BSA.

Stop Solution

95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF.

Other Enzymes

Incubation buffers were generally supplied by the manufacturer.

Bacteriophage T4 DNA Ligase

Bacteriophage T4 DNA ligase (New England Biolabs) was incubated in 50 mM Tris/HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 µg ml^{-1} BSA.

Bacteriophage T4 Polynucleotide Kinase

Bacteriophage T4 polynucleotide kinase (Boehringer Mannheim) was incubated in 50 mM Tris/HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine HCl, 0.1 mM EDTA.

Calf Intestinal Alkaline Phosphatase (CIP)

Calf intestinal alkaline phosphatase (CIP) (Boehringer Mannheim) was incubated in 1 mM $ZnCl_2$, 1 mM $MgCl_2$, 10 mM Tris/HCl (pH 8.3).

Dam Methylase

Dam methylase (New England Biolabs) was incubated in 50 mM Tris/HCl (pH 7.5), 10 mM EDTA, 1 mM DTT, 80 µM S-adenosylmethionine (SAM)

Klenow Enzyme

Klenow enzyme (Boehringer Mannheim) was incubated in restriction endonuclease buffer, as it was generally used subsequent to a restriction digest.

Other Solutions

dNTP Stocks

dNTPs (Sigma Chemical Company) were prepared in sterile distilled water at concentration of 50 mM and 2 mM. Stored at –20°C.

BSA (20 mg ml^{-1})

Bovine serum albumin (Boehringer Mannheim) was stored at –20°C.

Solutions for Gel Electrophoresis

Agarose Gel Electrophoresis

TAE Gel Buffer Stock (20×)

0.8 M Tris/acetate, 20 mM EDTA (pH 8.0).

TAE Gel-Loading Sample Buffer Stock (5×)

0.2 M Tris/acetate, 0.25 M EDTA (pH 8.0), 0.2% bromophenol blue, 15% Ficoll 400.

Ethidium Bromide (10 mg ml^{-1})

Prepared in sterile 1× TE buffer, protected from light at 4°C. Used at 0.5 μg ml^{-1}.

Polyacrylamide Gel Electrophoresis

TBE Gel Buffer Stock (10×)

0.89 M Tris/borate, 20 mM EDTA (pH 8.0).

Formamide-EDTA Gel-Loading Sample Buffer

98% formamide, 0.1% bromophenol blue, 0.1% xylene cyanol FF, 10 mM EDTA (pH 8.0).

40% Acrylamide Stock

38% (w/v) acrylamide, 2% (w/v) N-N-methylene bis-acrylamide (Northumbria Biologicals), protected from light at 4°C.

1× TBE 6% Acrylamide Solution

42.5 g urea, 10 ml 10× TBE buffer, 15 ml 40% acrylamide stock, distilled water to 100 ml total. Degassed and stored at 4°C in foil-wrapped bottle.

0.5× TBE 6% Acrylamide Solution for DNA Sequencing Gels

59.5 g urea, 7 ml 10× TBE buffer, 21 ml 40% acrylamide stock, distilled water to 140 ml total. Degassed and stored at 4°C in foil-wrapped bottle.

5× TBE 6% Acrylamide Solution for DNA Sequencing Gels

12.75 g urea, 15 ml 10× TBE buffer, 4.5 ml 40% acrylamide stock, 6 g sucrose, and a few grains of bromophenol blue. Degassed and stored at 4°C in foil-wrapped bottle.

TEMED

TEMED (N-N-N'-N'-tetra-methyl-1,2-diamino-ethane) (Sigma Chemical Company) was stored at 4°C and protected from light.

10% AMPS

10% AMPS (ammonium persulphate) was freshly prepared in distilled water.

Gel Fix

10% (v/v) glacial acetic acid, 10% (v/v) methanol.

Table 2.4

Restriction Endonucleases Used in this Work

The target site is shown in the 5'→3' direction. The point of cleavage is indicated by the down-pointing arrow (▼). Incubation temperature is 37°C unless otherwise indicated.

1. Blocked by Dam methylation (N^6-methyladenine).
2. Incubation temperature is 50°C.
3. Incubation temperature is 65°C.
4. Requires Dam methylation (N^6-methyladenine) for cleavage.

Table 2.4

Enzyme	Target Site	Supplier	Buffer	Notes
*Acc*I	GT▼(A/C)(G/T)AC	Northumbria Biologicals	2	
*Bam*HI	G▼GATCC	Northumbria Biologicals	4	
*Bcl*I	T▼GATCA	Northumbria Biologicals	5	1, 2
*Bgl*II	A▼GATCT	Northumbria Biologicals	7	
*Bst*BI	TT▼CGAA	New England Biolabs	*4*	3
*Dpn*I	GA▼TC	New England Biolabs	*4*	4
*Eco*RI	G▼AATTC	Northumbria Biologicals	6	
*Hin*dIII	A▼AGCTT	Northumbria Biologicals	10	
*Mbo*I	▼GATC	New England Biolabs	*3*	1
*Pvu*I	CGAT▼CG	Northumbria Biologicals	6	
*Sac*I	GAGCT▼C	Northumbria Biologicals	2	
*Sau*3AI	▼GATC	New England Biolabs	*2*	
*Sca*I	AGT▼ACT	Northumbria Biologicals	6	
*Ssp*I	AAT▼ATT	Northumbria Biologicals	6	
*Stu*I	AGG▼CCT	Boehringer Mannheim	B	
*Taq*I	T▼CGA	Northumbria Biologicals	4	3
*Xba*I	T▼CTAGA	Boehringer Mannheim	H	
*Xho*II	Pu▼GATCPy	Boehringer Mannheim	L	

Table 2.5

Composition of Incubation Buffers for Restriction Endonucleases

The incubation buffers were supplied by the manufacturers as 10× stock solutions. The values shown in Table 2.5 are the 1× final concentrations (mM). The buffers were generally supplemented with BSA (bovine serum albumin) (Boehringer Mannheim) to a final concentration of 100 μg ml^{-1}.

Table 2.5

Supplier	**Northumbria Biologicals**						**New England Biolabs**			**Boehringer Mannheim**		
Buffer	2	4	5	6	7	10	*2*	*3*	*4*	B	L	H
Tris/acetate	33	-	-	-	-	-	-	-	20	-	-	-
Tris/HCl	-	10	10	50	-	50	10	50	-	10	10	50
Glycine-NaOH	-	-	-	-	20	-	-	-	-	-	-	-
Mg-acetate	10	-	-	-	-	-	-	-	10	-	-	-
$MgCl_2$	-	5	10	10	10	10	10	10	-	5	10	10
K-acetate	66	-	-	-	-	-	-	-	50	-	-	-
NaCl	-	100	50	100	200	50	50	100	-	100	-	100
DTE	-	-	-	-	-	-	-	-	-	-	1	1
DTT	0.5	-	1	1	-	1	1	1	1	-	-	-
β-Mercaptoethanol	-	1	-	-	7	-	-	-	-	1	-	-
pH @ 37°C	8.2	8.3	7.8	7.8	9.5	8.3	7.9	7.9	7.9	8.0	7.5	7.5

SECTION 2

METHODS

Microbiological Methods

Bacterial Methods

Storage of Bacteria

For short term storage, bacteria were kept at 4°C on L agar plates sealed with parafilm. Permanent stocks were prepared by adding 5 drops of sterile 100% glycerol to 1 ml of a stationary phase culture in an Eppendorf. This was sealed with parafilm and stored at –70°C.

Growth of Bacteria

Temporary stocks were generated by streaking permanent stocks to single colonies on L agar plates, which were incubated at 37°C overnight unless otherwise indicated. Overnight cultures were grown by inoculating a single colony into L broth and shaking at 37°C, unless otherwise indicated.

Test of UV Sensitivity

This method was used to test the sensitivity of *E. coli dam* strains to ultra-violet (UV) light (see Chapter 6). An overnight culture was diluted 40-fold in L broth and grown shaking at 37°C to OD_{650} = 0.5. The culture was then serially diluted in ice-cold L broth and 100 µl of the 10^{-3} and 10^{-4} dilutions were spread on two sets of fresh, dry L agar plates. One set of plates was incubated at 37°C overnight. The other set was exposed for 20 seconds to an UV lamp (Ultraviolet Products Incorporated) giving 1 J m^{-2} s^{-1} illumination, wrapped immediately in foil (to prevent DNA photolyase activation) and incubated at 37°C overnight. Colonies were counted and the strain showing ~2% survival was chosen (McGraw & Marinus, 1980).

Bacteriophage λ Methods

Preparation of Plating Cultures

Plating cultures of the appropriate bacterial strain were made by diluting an overnight culture 10-fold in L broth p/c. This was grown shaking at 37°C for 2 – 2½ hours, depending upon the viability of the strain, and diluted with an equal volume of TM buffer. Plating cultures were stored at 4°C for up to 3 days.

Titration of Bacteriophage λ Stocks

Accurate Titration

The phage stock was diluted serially in phage buffer, and 100 µl of each appropriate dilution was added to 250 µl of the plating culture and left to adsorb for 20 minutes at room temperature. 2.5 ml of molten BBL top agar at 46°C were then added and vortexed very briefly at low speed to homogenise the mixture. This was poured immediately onto the surface of a fresh, dried BBL plate. The plaques were counted after incubating the plates overnight at 37°C (unless otherwise indicated).

Spot Tests

If a rough estimate of phage titre was required, 250 µl of the plating culture was briefly vortexed with 2.5 ml of molten BBL top agar (at 46°C) and poured onto the surface of a fresh, dried BBL plate. After this had set, 2 µl or 10 µl aliquots of the phage dilutions were spotted onto the bacterial lawn and allowed to dry before incubating the plates overnight at 37°C (unless otherwise indicated). This technique is particularly useful for comparing the plating efficiencies of different phage strains such as in cloning, as up to 16 (10 µl) and 52 (2 µl) spots may be compared on a single plate.

Plaque Area Assay

Plate Pouring

BBL agar (plaque area) was freshly prepared using the exact amounts set out in the Materials section. Plates were poured at 46°C using exactly 40 ml of agar and allowed to set in large stacks (> 12 plates). The plates were used after three days, although the terminal two plates of each stack were discarded. These precautions were necessary because small variations between plates, which are primarily due to drying conditions, can markedly affect the size of plaques on a lawn.

Phage Plating

A single plaque of the appropriate phage was picked using a sterile Pasteur pipette into 1 ml phage buffer. The mixture was vortexed shortly with 10 µl chloroform (to lyse remaining host cells) and left for one hour at 4°C to allow diffusion of phage from the agar plug. It was then diluted serially and titrated accurately on N2364 plating cells. The volume of the phage suspension corresponding to 200 – 300 PFU (plaque forming units) was calculated from the titre (typically ~20 µl of a 10^{-2} dilution), added to 250 µl of a fresh N2364 plating culture and left to adsorb for 20 minutes at 37°C. These conditions were chosen to maximise preadsorption of the

phage. Exactly 2.5 ml of molten BBL top agar (plaque area) at 46°C were then added and vortexed very briefly at low speed to homogenise the mixture. This was poured immediately onto the surface of a BBL (plaque area) plate (prepared as described above), and incubated at 37°C overnight.

Plaque Area Quantification

The area, width, length and perimeter of each plaque were measured using a Quantimet 970 digital image analyser (Cambridge Instruments). The images of the plaques were presented to the image analyser via a Chalnicon video camera. Optimum illumination was provided by placing a lightbox (20 cm × 20 cm) at 45 cm beneath the plate. The intensity of the illumination was automatically adjusted by the image analyser such that the peak brightness of the plaque images was equivalent to a video signal of 1 volt. Field width was set at approximately 20 mm, and the image was digitised into 896 by 704 pixels with 256 grey levels per pixel. The analysis of the image was performed by a Quips routine "PLAQUE" written by Dr C. E. Jeffree (Science Faculty EM Facility, University of Edinburgh). The plaques were first detected by setting appropriate grey-level threshold values. Following detection, single-pixel noise was removed by means of an erode procedure followed by a dilate procedure to restore surviving objects to their original size. An editing loop was provided to enable deletion of spurious objects, or the manual separation of overlapping plaques. Measurements of the plaques were obtained based on the detected region.

Bacteriophage λ Cross

A fresh overnight culture of the host strain was diluted 100-fold in L broth p/c and grown to OD_{650} = 0.45 (~2 × 10^8 cells ml^{-1}). The culture was centrifuged at 4.5 krpm for 5 minutes in an MSE Centaur-2 bench centrifuge, resuspended in a equal volume TM buffer and starved shaking at 37°C for 30 minutes. 250 µl of the starved cell culture was mixed with 250 µl of a phage suspension containing the two λ strains and incubated at 37°C for 30 minutes. Both phage strains are generally added at a multiplicity of infection (MOI) of 5, but this may be varied according to the circumstances (see Chapter 5). The cell culture with the adsorbed phage was then diluted 100-fold with pre-warmed L broth and grown shaking at 37°C for 1½ hours. Chloroform was added to a final concentration of 0.2% and shaking was continued for a further 5 minutes to complete lysis. Cell debris was pelleted by centrifugation at 4.5 krpm for 5 minutes (MSE Centaur-2 bench centrifuge). The supernatant was stored at 4°C and titrated on permissive and restrictive indicator strains.

Preparation of Bacteriophage λ Stocks by Plate Lysates

The required phage were titrated and single plaques were picked using a sterile Pasteur pipette into 1 ml phage buffer. The mixture was vortexed shortly with 10 µl chloroform and left for one hour at 4°C to allow diffusion of phage from the agar plug. 100 µl, 200 µl and 300 µl volumes of the phage suspension were added to 250 µl of a fresh plating culture and left to adsorb for 20 minutes at room temperature. 2.5 ml of molten Seakem top agar at 46°C were added and vortexed very briefly at low speed to homogenise the mixture. This was poured immediately onto the surface of a very fresh (< 1 hour old) Seakem agar plate and incubated at 37°C for 6 – 8 hours until the lawn reached confluent lysis. 4 ml of TM buffer was poured onto the lawn and a sterile 10 ml glass pipette was used to homogenise the top agar. The homogenate was decanted and a further 2 ml TM buffer poured on the plate to remove the remaining top agar. This was added to the homogenate, which was shortly vortexed with 50 µl chloroform and left at 4°C overnight to allow diffusion of phage from the agar. The suspension was then vortexed once more and centrifuged at 4.5 krpm for 10 minutes in an MSE Centaur-2 bench centrifuge to remove agar and cell debris. The supernatant was titrated, and stored at 4°C after adding 20 µl chloroform.

Preparation of Bacteriophage λ Stocks by Liquid Lysates

This method was used when large quantities of phage were required, such as in the preparation of λ DNA for a Dam methylation assay. A fresh overnight culture was diluted 50-fold in pre-warmed L broth (liquid lysis), and grown shaking at 37°C. The flask was no more than one-eighth full, as efficient λ lysis is dependent upon good aeration. When the culture had reached OD_{650} = 0.5 (~2.5×10^8 cells ml^{-1}) after ~1¼ hours, phage were added at a multiplicity of infection (MOI) of 0.1. The OD_{650} of the culture at the time of infection and the MOI were varied where necessary (see Chapter 8); less vigorous phage strains such as DRL179 were generally added at an MOI of 0.1 when the culture had reached OD_{650} = 0.3 (Fig. 8.1). The flasks were returned to 37°C with vigorous shaking and the OD_{650} was monitored periodically. After 2 – 4 hours of growth, cell lysis occurs and the OD_{650} begins to fall (from ~2.5). However, it can take another 2 – 4 hours of growth to reach the minimum OD_{650} (0.4 ~ 0.7). When two successive OD_{650} readings were similar, indicating that the minimum had been reached, 1 ml of chloroform per 500 ml of lysate was added and shaking was continued for a further 10 minutes to complete lysis. The lysate was then centrifuged at 10 krpm for 15 minutes at 4°C (Sorvall centrifuge, GSA rotor) to remove cell debris. The supernatant was titrated and stored at 4°C.

Purification of Bacteriophage λ Particles from Plate Lysates

This method was used to purify λ particles from plate lysates in order to extract λ DNA by the small scale method. 1 ml of a plate lysate was transferred to an Eppendorf, taking care to avoid any chloroform. 5 µl DNase and 5 µl RNase A (both 10 mg ml^{-1}) were added and incubated at 37°C for 30 minutes to digest bacterial DNA and RNA. The phage suspension was centrifuged for 1 minute at 15 krpm (Sorvall Microspin 24 centrifuge) and the supernatant added to 0.5 ml ice-cold 30% PEG 6000, 3 M NaCl in a pre-cooled Eppendorf. The Eppendorf was inverted five times to mix the suspensions and incubated on ice for 2 – 4 hours. It was then centrifuged at 15 krpm for 15 minutes at 4°C (Sorvall Microspin 24 centrifuge) and the supernatant was removed carefully. The Eppendorf was centrifuged again for 30 seconds and any remaining supernatant removed. The pellet was gently resuspended in 200 µl TM buffer, 200 µl of chloroform was added and the Eppendorf vortexed very briefly three times. The chloroform causes PEG 6000 to precipitate, leaving the λ particles in suspension. The phases were separated and the PEG 6000 pelleted by centrifugation at 15 krpm for two minutes (Sorvall Microspin 24 centrifuge). The aqueous (upper) layer was transferred to a fresh Eppendorf and stored at 4°C.

Purification of Bacteriophage λ Particles from Liquid Lysates

This method was used to purify λ particles from liquid lysates, in order to extract λ DNA of high purity by the large scale method.

Precipitation of λ Particles with PEG 6000

DNase and RNase were added to the liquid lysate to a final concentration of 1 µg ml^{-1} and incubated at room temperature for 30 minutes, swirling occasionally. NaCl was then added to the liquid lysate to a final concentration of 1 M, dissolved gently and incubated for one hour on ice (or left at 4°C overnight). It was then centrifuged at 10 krpm for 10 minutes at 4°C (Sorvall centrifuge, GSA rotor) to remove bacterial debris that had been liberated by DNase and RNase digestion and the high NaCl concentration. The supernatant was decanted into a clean flask, PEG 6000 was added to a final concentration of 10% (w/v) and dissolved gently at room temperature. The suspension was incubated on ice for 4 – 6 hours and then centrifuged at 10 krpm for 15 minutes at 4°C (Sorvall centrifuge, GSA rotor). The supernatant was discarded and the centrifuge bottle allowed to drain for 10 minutes. The pellet was resuspended at room temperature in 5 ml phage buffer by gentle shaking for one hour. The suspension was transferred to a glass bottle and a further 1 ml phage buffer used to rinse the inside of the centrifuge bottle; this was added to the suspension. 7 ml of

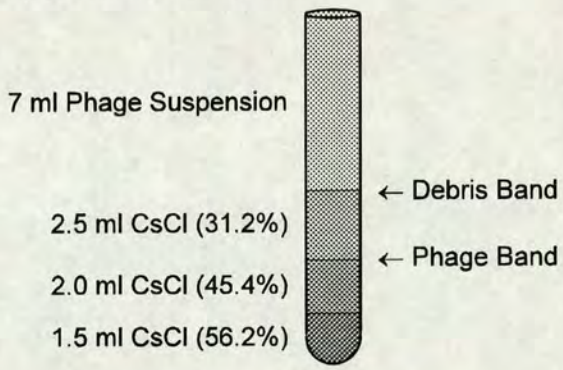
chloroform was added to the suspension and vortexed gently for 10 seconds. The phases were separated and the PEG 6000 pelleted by centrifugation at 4.5 krpm for 10 minutes (MSE Centaur-2 bench centrifuge). The aqueous layer was transferred to a fresh bottle and stored at 4°C.

Purification of λ Particles by CsCl Step Gradient Centrifugation

A CsCl step gradient was made by successively underlayering 2.5 ml of 31.2%, 2.0 ml of 45.4% and 1.5 ml of 56.2% CsCl solutions in a 13 ml Beckman Ultra-Clear™ tube using a Pasteur pipette. This is shown schematically in Fig. 2.1. The phage suspension was then carefully loaded on top of the CsCl solution layers to within 2 mm of the top of the tube. The step gradients were balanced to within ±10 mg and centrifuged at 35 krpm for 35 minutes at 20°C (Sorvall ultracentrifuge, TH641 titanium swing-out rotor). Two bands could be seen in visible light: the lower was phage particles and the upper was debris (Fig. 2.1). The phage band (~1 ml) was collected through the side of the tube using a hypodermic needle (19G1½ 1.1 × 40 mm) and syringe.

Figure 2.1

CsCl step gradient for purification of λ particles.



The band from the step gradient was dialysed against 2 *l* of phage buffer for 2 hours at 4°C. The dialysed suspension was then loaded onto a second CsCl step gradient, which had been prepared as before. In order to minimise phage loss, the inside of the dialysis tubing was rinsed with 1 ml phage buffer; this was also loaded onto the CsCl step gradient. The tubes were balanced as before and centrifuged at 35 krpm for 35 minutes at 20°C (Sorvall ultracentrifuge, TH641 titanium swing-out rotor). Normally, only the lower (phage) band was visible after the second round of centrifugation. This was collected as before using a hypodermic needle and syringe, and stored at 4°C.

In Vitro Packaging of Bacteriophage λ DNA

Buffer A and buffer M1 were thawed at room temperature, while sonicated extract and freeze-thaw lysate were thawed on ice. The components were mixed in this order:

1. Buffer A 7 µl
2. λ DNA 5 µl
3. Buffer M1 2 µl
4. Sonicated extract 10 µl
5. Freeze-thaw lysate 10 µl

The reaction was incubated at 25°C for 1 hour, then diluted with 500 µl of phage buffer. 100 µl, 10 µl and 1 µl aliquots of this phage suspension were titrated.

Plasmid Methods

Maintenance of Plasmids

Plasmids were maintained by supplementing the media with antibiotics (Table 2.3).

Transformation

This method is derived from Mandel & Higa (1970).

Preparation of Competent Cells

A fresh overnight culture was diluted 100-fold in L broth and grown shaking at 37°C to an OD_{650} = 0.4. The culture was chilled on ice for 15 minutes, and 20 ml were centrifuged at 4.5 krpm for 5 minutes at 4°C (MSE Centaur-2 bench centrifuge). The cell pellet was resuspended in 2 ml ice-cold 0.1 M $CaCl_2$ and incubated on ice for 20 minutes. The suspension was again centrifuged at 4.5 krpm for 5 minutes at 4°C (MSE Centaur-2 bench centrifuge), and resuspended gently in 400 µl of ice-cold $CaCl_2$. The competent cells were used immediately.

Transformation

A 200 µl aliquot of competent cells was added to 50 – 100 ng plasmid DNA (suspended in a minimum volume of 1× TE buffer or distilled water) and incubated on ice for 30 minutes. The cells were heat-shocked for 2 minutes in a 42°C waterbath and rapidly cooled on ice. 10 µl and 100 µl aliquots of the transformed cells were spread on L agar plates containing the appropriate antibiotic and incubated overnight at 37°C unless indicated otherwise.

Detection of Cloned Inserts in pMS2B

L agar was supplemented with Xgal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) at a final concentration of 40 µg ml^{-1} and ampicillin at a final concentration of 100 µg ml^{-1}. Plasmids with inserts at the *Acc*I site gave rise to white colonies while those without gave rise to blue colonies (see Chapter 11).

DNA Purification and Manipulation

General Methods of DNA Purification

Phenol, Phenol-Chloroform and Chloroform Extraction

Successive phenol-chloroform and chloroform extractions were generally sufficient to purify DNA from contaminating protein. An initial phenol extraction was used where the purity of DNA was paramount, or where the purpose of the phenol was to disrupt λ phage heads (see large scale method for purification of λ DNA).

Phenol Extraction

Phenol extraction was carried out in Eppendorf tubes. An equal volume of equilibrated phenol was added to the DNA solution and the two were homogenised by vortexing briefly. The phases were separated by centrifugation at 15 krpm for 5 minutes (Sorvall Microspin 24 centrifuge). The aqueous (upper) layer was transferred to a fresh Eppendorf.

Phenol-Chloroform Extraction

Phenol-chloroform extraction was carried out identically to phenol extraction, except that equilibrated 25:24:1 phenol-chloroform-isoamylalcohol was used in place of phenol.

Chloroform Extraction

Chloroform extraction was carried out identically to phenol extraction, except that 24:1 chloroform-isoamylalcohol was used in place of phenol, and the phases were separated by centrifugation at 15 krpm for only one minute.

Ethanol and Isopropanol Precipitation

Ethanol precipitation was normally used to concentrate and purify DNA. Isopropanol precipitation was used where a large volume was impractical, or where it was desirable to avoid the coprecipitation of RNA.

Ethanol Precipitation

Ethanol precipitation was carried out in 15 ml Corex tubes or Eppendorf tubes. 0.1 volumes of 3 M sodium acetate pH 5.3 were added to 0.9 volumes of DNA solution, followed by 2 volumes freezer-cold (–20°C) ethanol. Where the concentration of EDTA in the DNA solution exceeded 10 – 15 mM, 3 M sodium acetate pH 7.0 was used; high concentrations of EDTA precipitate from solution at acid pH. The tube contents were mixed by inversion and incubated at –20°C for at least one hour to

allow precipitation. The DNA was then pelleted by centrifugation at 15 krpm for 20 minutes at 4°C (Sorvall centrifuge, SS34 rotor, or Sorvall Microspin 24 centrifuge). The ethanol was discarded, the tube filled with freezer-cold 70% ethanol, and recentrifuged at 15 krpm for 10 minutes at 4°C. The supernatant was discarded, the pellet was air dried for 15 minutes and redissolved in 1× TE buffer or distilled water.

Isopropanol Precipitation

Isopropanol precipitation was carried out identically to ethanol precipitation, except that 0.7 volumes of isopropanol (at room temperature) were used in place of 2 volumes freezer-cold ethanol The tube was incubated at room temperature for 15 minutes to allow precipitation, and centrifugation (15 krpm for 20 minutes, as above) was also carried out at room temperature. The pellet was rinsed with freezer-cold 70% ethanol and recentrifuged as described above.

Purification of DNA from Agarose Gels and Solutions

Purification of DNA from agarose gels and solutions was performed using GENECLEAN® and MERmaid™ kits (BIO 101). The procedures are based on the method of Vogelstein & Gillespie (1979).

GENECLEAN® Procedure

The band of interest was excised from the agarose gel and its weight determined in a tared Eppendorf tube. 3 volumes (v/w) of 6 M NaI stock were added and the agarose gel dissolved by 5 minutes incubation in a 50°C waterbath, with occasional mixing by tube inversion. Where the DNA was in aqueous solution, 3 volumes (v/v) of 6 M NaI stock were added without incubation at 50°C. 5 µl of GLASSMILK® suspension were added, mixed by vortexing and incubated on ice for 5 minutes, vortexing once every minute. The silica matrix with the bound DNA was pelleted by centrifugation for 15 seconds at 15 krpm (Sorvall Microspin 24 centrifuge) and the supernatant discarded. The pellet was resuspended in 500 µl freezer-cold NEW wash and centrifuged for 15 seconds at 15 krpm (Sorvall Microspin 24 centrifuge). The supernatant was discarded and the pellet was washed a further two times with 500 µl NEW wash. After the supernatant from the third wash had been removed, the tube was centrifuged again for 15 seconds at 15 krpm and the remaining liquid removed. The pellet was air dried for 10 minutes, resuspended in 5 – 10 µl 1× TE buffer or distilled water and incubated in a 50°C waterbath for 3 minutes. It was then centrifuged for 1 minute at 15 krpm (Sorvall Microspin 24 centrifuge) and the supernatant containing the eluted DNA was transferred to a fresh Eppendorf tube. A second elution was identically performed, and the supernatants were combined.

MERmaid™ Procedure

This method was only used to purify DNA from aqueous solutions. 3 volumes (v/v) of high salt binding solution were added to the DNA solution. 8 µl of GLASSFOG™ suspension were added, mixed by vortexing and incubated at room temperature for 15 minutes, vortexing once every few minutes. The silica matrix with the bound DNA was pelleted by centrifugation for 15 seconds at 15 krpm (Sorvall Microspin 24 centrifuge) and the supernatant discarded. The pellet was resuspended in 300 µl ethanol wash by vortexing and centrifuged for 15 seconds at 15 krpm (Sorvall Microspin 24 centrifuge). The supernatant was discarded and the pellet was washed one more time with 300 µl ethanol wash. The tube was centrifuged again for 15 seconds at 15 krpm and any remaining liquid removed. The pellet was air dried for 10 minutes, resuspended in 5 – 10 µl 1× TE buffer or distilled water and incubated at room temperature for 5 minutes. It was then centrifuged for 2 minutes at 15 krpm (Sorvall Microspin 24 centrifuge) and the supernatant containing the eluted DNA was transferred to a fresh Eppendorf tube. A second elution was identically performed, and the supernatants were combined.

Bacteriophage λ DNA Purification

Small Scale Method

This method was used to purify λ DNA from plate lysates. The λ particles were purified from plate lysate phage stocks as described earlier and resuspended in 200 µl TM buffer (see: Purification of Bacteriophage λ Particles from Plate Lysates). The λ suspension was extracted with 200 µl of phenol to disrupt the phage heads, and then extracted with 200 µl of phenol-chloroform and chloroform to purify the λ DNA from denatured capsid proteins. The DNA was precipitated with ethanol, rinsed with 70% ethanol and resuspended in 36 µl 1× TE buffer. The DNA was then precipitated with isopropanol and rinsed with 70% ethanol. The λ DNA pellet was resuspended in 20 µl 1× TE buffer and stored at –20°C.

Large Scale Method

This method was used to purify λ DNA from liquid lysates. The λ particles were purified from liquid lysate phage stocks as described earlier (see: Purification of Bacteriophage λ Particles from Liquid Lysates). The phage band from the second round of CsCl step gradient centrifugation was dialysed overnight against 2 *l* of 1× TE buffer at 4°C. Pronase was added to the phage band (in the dialysis tubing) to a final concentration of 1 mg ml^{-1}. This was then dialysed for 2 hours against 1 *l* of

pronase dialysis buffer at 37°C. The phage DNA solution was transferred to Eppendorf tubes and the inside of the dialysis tubing was rinsed with 500 µl 1× TE buffer to minimise loss of phage DNA. The pooled DNA solutions were extracted with an equal volume of phenol to disrupt remaining phage heads, and then extracted with equal volumes of phenol-chloroform and chloroform to purify the λ DNA from denatured protein. The DNA was precipitated with ethanol, rinsed with 70% ethanol and resuspended in a total of 180 µl 1× TE buffer. The DNA was then precipitated with isopropanol and rinsed with 70% ethanol. The λ DNA pellet was resuspended in 100 µl 1× TE buffer and stored at –20°C. The DNA concentration was estimated using its absorbance at 260 nm as determined by a Perkin-Elmer Lambda 15 spectrophotometer.

Plasmid DNA Purification

Small Scale Method

This method is derived from Birnboim & Doly (1979). 1.5 ml of a fresh overnight culture were transferred to an Eppendorf tube and centrifuged for 30 seconds at 15 krpm (Sorvall Microspin 24 centrifuge). The supernatant was discarded and the tube centrifuged again for 5 seconds at 15 krpm. The remaining liquid was removed and the pellet was resuspended in 100 µl ice-cold TEG by vortexing two Eppendorf tubes simultaneously for 15 seconds. The cell suspension was incubated at room temperature for 5 minutes. 200 µl of freshly prepared alkaline SDS were added, the tube was inverted five times to mix the solutions and incubated for 5 minutes on ice. 150 µl ice-cold potassium acetate were added and the tube was vortexed upright for five seconds. After 5 minutes incubation on ice, the cell debris was removed by centrifugation at 15 krpm for 5 minutes at 4°C (Sorvall Microspin 24 centrifuge). 400 µl of supernatant containing the plasmid DNA were transferred to a fresh Eppendorf tube and extracted with equal volumes of phenol-chloroform and chloroform. Two volumes of 100% ethanol (at room temperature) were added, the tube was inverted to mix the components and incubated at room temperature for 5 minutes. The DNA was pelleted by centrifugation at 15 krpm for 10 minutes at 4°C (Sorvall Microspin 24 centrifuge). The supernatant was discarded and the pellet was rinsed with 1 ml freezer-cold 70% ethanol. After further centrifugation at 15 krpm for 5 minutes at 4°C, the 70% ethanol was removed and the DNA pellet was air dried for 15 minutes. The plasmid DNA was redissolved in 50 µl TE-RNase A and stored at –20°C.

Large Scale Method

The large scale purification of plasmid DNA was performed using a QIAGEN plasmid midi kit. A fresh overnight culture was centrifuged at 10 krpm for 10 minutes at 4°C (Sorvall centrifuge, GSA or SS34 rotor). The volume of the overnight culture was a function of the plasmid type: low-copy vectors derived from pBR322 (e.g. pZTA1) were prepared from 100 – 250 ml cultures, while high-copy vectors derived from pUC18 (e.g. pMS2B) were prepared from 25 – 50 ml cultures. The cell pellet was resuspended in 4 ml buffer P1 and transferred to a 30 ml polypropylene centrifuge tube. 4 ml of buffer P2 were added, mixed gently by inversion, and incubated at room temperature for 5 minutes. 4 ml of ice-cold buffer P3 were added, the solution was mixed by inverting the tube 5 times, and incubated on ice for 15 minutes. The sample was mixed once more and centrifuged at 16 krpm for 30 minutes at 4°C (Sorvall centrifuge, SS34 rotor). The supernatant was promptly removed and applied to a QIAGEN-tip 100 which had previously been equilibrated using 4 ml buffer QBT. After the column had emptied by gravity flow, it was washed by two applications of buffer QC. This removes contaminants in the DNA preparation from the QIAGEN-tip 100. After the wash buffer had completely emptied, the DNA was eluted from the column with 5 ml buffer QF and allowed to empty into a 15 ml Corex tube. 3.5 ml (0.7 volumes) of isopropanol (at room temperature) were added to the DNA solution and the two were mixed by inversion. The mixture was then centrifuged at 12 krpm for 30 minutes at 4°C (Sorvall centrifuge, SS34 rotor). The supernatant was discarded and the DNA pellet was washed in freezer-cold 70% ethanol. After further centrifugation at 12 krpm for 10 minutes at 4°C, the 70% ethanol was removed and the DNA pellet was air dried for 15 minutes. The plasmid DNA was redissolved in 200 µl 1× TE buffer.

DNA Manipulation

Annealing of Oligonucleotides

As the majority of oligonucleotides used in this work were palindromic, special precautions were taken to promote the formation of double-stranded DNA over self-annealed hairpins. Around 40 µg of the oligonucleotide (Oswel DNA Service) was annealed in 1× TE buffer that had been made 10 mM NaCl. This was prepared by adding 40 µl of an aqueous solution of the oligonucleotide to 8 µl of 10× TE buffer and 32 µl of 25 mM NaCl in an Eppendorf tube. The mixture was heated to 100°C in a waterbath containing a large volume (> 500 ml) of water. The waterbath was

switched off and the oligonucleotide allowed to cool to room temperature gradually. It was then diluted 100-fold in 1× TE buffer 10 mM NaCl and stored at –20°C.

Restriction Digests

The restriction endonucleases used in this work are shown in Table 2.4. Restriction digests were carried out in 10 – 40 µl of the appropriate incubation buffer as shown in Table 2.5. The buffers were supplied as 10× stock solutions, and 0.1 volumes were added to the DNA solution to prepare the incubation buffer. This was supplemented by BSA (bovine serum albumin) (Boehringer Mannheim) to a final concentration of 100 µg ml^{-1}. A five-fold excess of restriction endonuclease (5 U μg^{-1} DNA) was added; this had been experimentally determined to be the optimum ratio of enzyme to DNA. The sample was then incubated for two hours at 37°C unless otherwise indicated (see Table 2.4). Where necessary, the reaction was stopped by phenol-chloroform and chloroform extractions, or heating to 65°C for 10 minutes. The latter method was used in the case of λ DNA restriction digests to denature the cohesive (*cos*) ends. However, this method was not used where the palindrome had been excised by an *Eco*RI digest, as the heating may promote hairpin formation and the loss of the palindromic DNA fragment (see Chapter 6).

DNA Sequencing

DNA sequencing was performed using a Sequenase® v2.0 DNA sequencing kit (United State Biochemical). The method is derived from Sanger *et al.* (1977) and uses dideoxynucleotide chain-terminating reactions. 5 – 10 µg λ DNA or 2 – 5 µg plasmid DNA were purified from aqueous solution using the GENECLEAN® procedure, and resuspended in 9 µl distilled water. 1 µl (~0.5 µg) sequencing primer (Oswel DNA Service) was added to the DNA in an Eppendorf tube. The DNA was then denatured by incubation in a waterbath at 100°C for 4 minutes, and rapid cooling on dry ice/ethanol. The labelling reaction shown below was prepared in a Eppendorf tube:

1. DTT 2 µl
2. [α^{35}S]dATP (Amersham) 0.5 µl (5 µCi)
3. Diluted labelling mix (1× in distilled water) 0.7 µl
4. Sequenase® buffer (5×) 2 µl
5. Diluted Sequenase® enzyme (1/8 in ice-cold enzyme dilution buffer) 2.5 µl

The labelling reaction was kept on ice while the template-primer mix was thawed and immediately centrifuged at 15 krpm for 10 seconds to remove the DNA solution from the walls of the tube. The template-primer mix was then added to the labelling

reaction and incubated at room temperature for 5 minutes. 4 µl aliquots of the labelling reaction were then transferred to four Eppendorf tubes, each containing 2 µl of either ddG, ddA, ddT or ddC termination mix; the tubes had been pre-warmed at 37°C (or 42°C) for 1 minute before the addition of 4 µl labelling reaction. The termination reactions were then incubated at 37°C for 3 minutes. Where the template was capable of forming secondary structures (and thereby eliciting band compression) the termination reactions were performed at 42 – 45°C. The reactions were halted by adding 4 µl stop solution to each tube and placing the sample on ice. The sequencing reactions were stored at –20°C until gel electrophoresis.

DNA Ligation

DNA ligation was carried out in a total volume of 10 – 20 µl bacteriophage T4 DNA ligase buffer. This was supplied as a 10× stock of which 0.1 volumes were added to the DNA solution. 40 units of bacteriophage T4 DNA ligase (New England Biolabs) were then added and the reaction was incubated overnight at 15°C. The reaction was terminated by heating to 70°C; the solution was allowed to cool to room temperature slowly to promote the reassociation of double-stranded DNA. DNA fragments were generally ligated with the vector molecule at a 3 – 5-fold molar ratio of cohesive ends.

DNA Phosphorylation

DNA phosphorylation was carried out in a total volume of 50 µl bacteriophage T4 polynucleotide kinase buffer. This was supplied as a 10× stock of which 0.1 volumes were added to an aqueous solution of the oligonucleotide containing ~30 pmol of 5' DNA termini. 50 pmol of ATP and 18 units of bacteriophage T4 polynucleotide kinase (Boehringer Mannheim) were then added, and the reaction was incubated at 37°C for 30 minutes. The reaction was terminated by the addition of 2 µl 0.5 M EDTA, phenol-chloroform and chloroform extraction, and ethanol precipitation using sodium acetate (pH 7.0). The DNA was resuspended in 1× TE buffer 10 mM NaCl for annealing of oligonucleotides (described earlier).

Dephosphorylation of DNA

Dephosphorylation of DNA was carried out in a total volume of 50 µl calf intestinal alkaline phosphatase (CIP) buffer. This was supplied as a 10× stock of which 0.1 volumes were added to an aqueous solution containing 1 µg DNA (~0.625 pmol 5' termini). 1 unit of CIP (Boehringer Mannheim) was then added, and the reaction was incubated at 37°C for 30 minutes. A second aliquot (1 unit) of CIP was added and the

incubation continued at 37°C for a further 30 minutes. The reaction was terminated by the addition of 1 µl 0.5 M EDTA and heating to 75°C for 10 minutes. The DNA was purified by phenol-chloroform and chloroform extraction, and ethanol precipitation using sodium acetate (pH 7.0). The DNA was resuspended in 1× TE buffer.

In Vitro Dam Methylation

In vitro Dam methylation was carried out in a total volume of 60 µl Dam methylase buffer. This was supplied as a 10× stock of which 0.1 volumes were added to an aqueous solution of 6 µg λ DNA in an Eppendorf tube. The S-adenosylmethionine (SAM) was provided separately and added to the 1× Dam methylase buffer at a final concentration of 80 µM. The Eppendorf tube containing the λ DNA in 1× Dam methylase buffer was pre-warmed to 37°C and 36 units of Dam methylase (New England Biolabs) were added; this is a ratio of 6 U Dam methylase per µg λ DNA. The reaction was incubated at 37°C. At predetermined times (such as 0, 10 and 60 minutes), 20 µl aliquots were transferred to pre-cooled Eppendorf tubes and the reaction stopped by placing the samples on dry ice (see Chapter 6). The samples were thawed on ice and the DNA purified by phenol-chloroform and chloroform extraction, and ethanol precipitation. The DNA was resuspended in 1× TE buffer.

Radiolabelling of DNA

Radiolabelling of DNA was generally carried out directly after a restriction digest. As Klenow enzyme functions adequately in virtually all restriction endonuclease buffers, the incubation buffer was not changed. 1 µl (10 µCi) of [α^{32}P]dATP or [α^{35}S]dATP (Amersham) was added to the restriction digest. This was supplemented by 1 µl of 2 mM stock solutions of appropriate non-radioactive dNTPs, as determined by the composition of the 5' overhang produced by the restriction endonuclease. For example, after *Eco*RI cleavage (shown below), non-radioactive dTTP was used.

```
5' ~G      AATTC~ 3'
3' ~CTTAA      G~ 5'
```

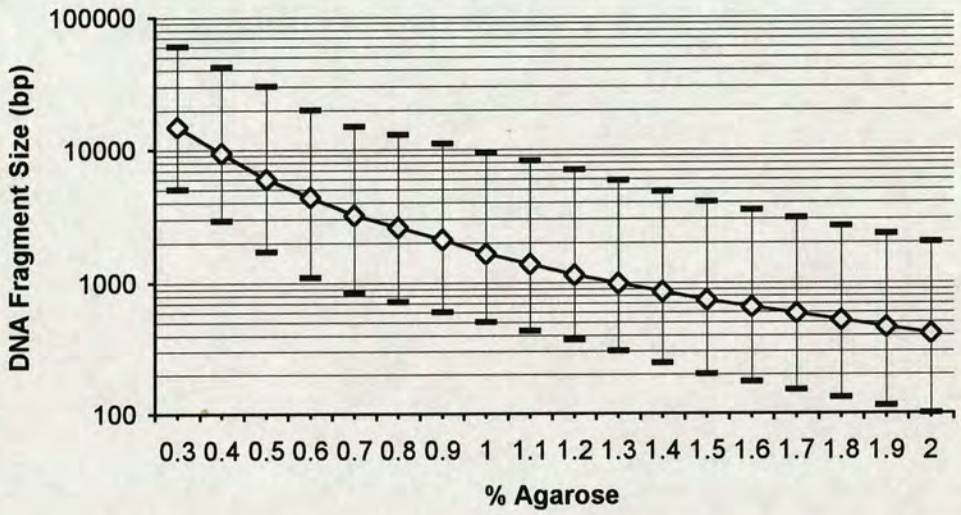
1 µl Klenow enzyme (1 U) was added and the reaction was incubated at room temperature for 15 minutes. A further 1 µl Klenow enzyme was added and the incubation continued for a further 10 minutes at room temperature. 1 µl of 50 mM stock solutions of dGTP, dATP, dCTP and dTTP were added as a chase (to ensure flush termini were produced) and the incubation continued for a further 5 minutes at room temperature. The DNA was purified by phenol-chloroform and chloroform extractions and ethanol precipitation, and resuspended in 1× TE buffer.

Gel Electrophoresis

Agarose Gel Electrophoresis

Agarose gels were made and run in 1× TAE gel buffer. The concentration of agarose was a function of the size of relevant DNA fragments, and varied from 0.5 to 2%. An agarose concentration of 0.7 – 0.8% was generally used to resolve fragments of divergent sizes. The optimal concentration of agarose is shown in Fig. 2.2.

Figure 2.2
Resolution of agarose gel electrophoresis (DNA fragment size vs. optimal % agarose).



To prepare an agarose gel, the appropriate amount of Seakem agarose (FMC Bioproducts) was dissolved in 100 ml 1× TAE buffer by simmering in a microwave oven. The molten agarose was cooled to 46°C and ethidium bromide was added to a final concentration of 0.5 μg ml^{-1}. The agarose was poured into a 14 cm × 11 cm perspex mould with a 14-tooth comb and allowed to set. The gel was then placed into a BRL model H5 horizontal gel electrophoresis tank containing 1 *l* 1× TAE buffer with 0.5 μg ml^{-1} ethidium bromide and the comb was removed. If DNA purification from the agarose gel was intended, ethidium bromide was added to neither the gel nor the running buffer; instead, the gel was stained after electrophoresis for 30 minutes in 0.5 μg ml^{-1} ethidium bromide. DNA samples were prepared by the addition of 0.2 volumes of TAE gel-loading sample buffer, and loaded into the gel slots. Gels were generally run overnight at ~1 V cm^{-1} (~26 V) and examined by 304 nm UV light on a C-62 Blak-Ray transilluminator (Ultraviolet Products Incorporated). Photographs were taken on Polaroid 667 film using a Wratten 25 red filter with an exposure of 1 second and an aperture of f/5.6½. DNA fragments for purification were excised using a sterile scalpel blade.

Polyacrylamide Gel Electrophoresis

Polyacrylamide gels (40 cm × 21 cm × 0.4 mm) were made and run in TBE gel buffer using a Sequi-Gen® Nucleic Acid Sequencing Cell (BioRad). The apparatus was prepared according to the manufacturer's instructions.

6% Denaturing Polyacrylamide Gels

The glass plates were assembled using 0.4 mm spacers and clamps provided by the manufacturer (BioRad). 60 µl TEMED and 150 µl freshly prepared 10% AMPS were added to 12 ml 1× TBE 6% acrylamide solution, and the mixture was promptly used to impregnate a 25 cm × 5 cm strip of blotting paper in a casting tray. The bottom edge of the glass plate sandwich was pushed firmly against the blotting paper and the catalysed acrylamide was allowed to enter the mould by capillary action. After this had set (~2 minutes), thereby sealing the bottom edge, the glass plate sandwich was laid at an angle of ~20° to the horizontal. 40 µl TEMED and 100 µl 10% AMPS were added to 35 ml 1× TBE 6% acrylamide solution, and the mixture was poured into the gel mould using a 25 ml glass pipette. A 16-tooth comb was inserted to produce flush wells of 0.9 mm width. The top of the mould was covered in Saran wrap and left overnight at 4°C. The gel mould was then assembled in the apparatus according to the manufacturer's instructions. 350 ml 1× TBE buffer were poured into the cathodic buffer chamber, the gel mould was clamped into place, and the anodic buffer chamber was filled with 1× TBE buffer. The comb was removed and the gel was run at a constant power of 40 W until a temperature of 50 – 55°C had been attained. DNA samples were prepared by resuspending the ethanol-precipitated DNA pellets in 10 µl formamide-EDTA gel-loading sample buffer. Samples were denatured by boiling in a waterbath for 3 minutes and cooling rapidly on ice. 5 µl of the sample were then loaded onto the gel using a micropipetter, and the gel was run at a constant power of 38 W for 3¾ hours. After electrophoresis, the gel mould was dismantled and the glass plate with the polyacrylamide gel placed in 2½ *l* gel fix for 30 minutes. The gel was transferred to wet blotting paper (Ford Goldmedal), covered in Saran wrap, and dried in a BioRad Model 583 Gel Drier for 1 hour at 80°C.

DNA Sequencing Gels

DNA sequencing gels consisted of 0.5 – 5× TBE buffer gradient 6% polyacrylamide denaturing gels. The glass plates were assembled as described above, except that the bottom edge of the mould was sealed using 0.5× TBE 6% acrylamide solution instead of 1× TBE 6% acrylamide solution. TEMED and 10% AMPS were added to 0.5× TBE 6% acrylamide and 5× TBE 6% acrylamide solutions as shown in Table 2.6:

Table 2.6

Volumes of acrylamide solution, TEMED and 10% AMPS for DNA sequencing gels.

	Acrylamide Solution	TEMED	10% AMPS
0.5× TBE 6% Acrylamide Mix	25 ml	25 µl	70 µl
5× TBE 6% Acrylamide Mix	8 ml	6 µl	16 µl

The gradient gel was poured using a 25 ml glass pipette: the 0.5× TBE 6% acrylamide mix and 5× TBE 6% acrylamide mix were taken up in the same pipette (in the order stated), and poured carefully into the gel mould. The flush side of a 24-tooth shark's-tooth comb was inserted, the top of the mould was covered in Saran wrap, and left overnight at 4°C. The gel mould was then assembled in the apparatus and pre-run to a temperature of 50°C as described above. The shark's-tooth comb was reinserted with the toothed side first. The DNA sequencing reactions were denatured by boiling in a waterbath for 3 minutes and cooling rapidly on ice. 5 µl of each reaction were then loaded onto the gel in the order: ddG, ddA, ddT, ddC. The gel was run at a constant power of 38 W for times ranging between 2 and 5 hours. After electrophoresis, the gel was fixed and dried as described above.

Autoradiography

^{32}P Isotopes

Autoradiography was carried out at –70°C in Cronex (DuPont) cassettes containing "xtra-life" intensifying screens, using Cronex 4 (DuPont) X-ray film (30 cm × 40 cm) preflashed to an OD_{540} of 0.1. The exposure time was varied from 4 – 16 hours, and films were developed in an X-OGRAPH Compact X2 automatic film processor.

^{35}S Isotopes

Autoradiography was carried out at room temperature in Cronex (DuPont) cassettes, using Cronex 4 (DuPont) X-ray film (30 cm × 40 cm). The exposure time was varied from 1 – 7 days, and films were developed in an X-OGRAPH Compact X2 automatic film processor.

Densitometry

Band intensities on autoradiographs were measured using a Shimadzu CS-930 dual-wavelength Chromato Scanner and DR-2 data recorder. The slit size was 6 × 0.7 mm, with a wavelength of 550 nm. Where necessary, several exposures of the film were taken to ensure that the response was approximately linear.

Quantification by PhosphorImager™

Dried polyacrylamide gels containing ^{35}S isotope were exposed to storage phosphor screens (Molecular Dynamics) for 1 – 3 days. The screens were scanned using a Molecular Dynamics Series 400 or Series 425 PhosphorImager™. Band intensities were measured using ImageQuant™ v3.3 software (Molecular Dynamics). A rectangular object boundary approximating the width of the relevant lane was delineated, and quantification performed using the default settings. The curve smoothing parameter was altered where necessary to allow the quantification of faint bands.

RESULTS

CHAPTER 3

*Eco*RI METHYLASE AS A PROBE OF CRUCIFORM EXTRUSION *IN VIVO*

Introduction

As described in the Introduction to this thesis, DNA methyltransferases are admirable probes of DNA structure. A number of different methylases have been used both *in vitro* and *in vivo* (Zacharias, 1992), but arguably the commonest choice of enzyme is *Eco*RI methylase (M*Eco*RI); most significantly, it has been employed in the detection of left-handed Z-DNA *in vivo* (Jaworski *et al.*, 1987). The use of M*Eco*RI as a probe of cruciform extrusion *in vivo* was therefore considered and investigated.

The procedure employed by Jaworski *et al.* (1987) utilises a temperature-sensitive *Eco*RI methylase (M*Eco*RI^{ts}) to detect Z-DNA formation by $(C\text{-}G)_n$ sequences in *E. coli*. The use of an inducible methylase probe was necessary as all target sites (including those which have a potential for Z-DNA formation) were readily methylated when the wild-type M*Eco*RI was used. The authors concluded that if the M*Eco*RI gene is permanently expressed, target sites will be methylated while the plasmid is replicating and not yet supercoiled (Lyons & Schendel, 1984). As Z-DNA formation is thermodynamically unfavourable in relaxed DNA, the target site located in the $(C\text{-}G)_n$ sequence would be methylated normally. In order to ensure that the plasmid containing the $(C\text{-}G)_n$ sequence is replicated and supercoiled prior to activating the probe, the authors controlled the expression of an inducible methylase using the M*Eco*RI^{ts} gene. In reducing the overmethylation that arises through the use of a constitutively expressed enzyme, Jaworski *et al.* (1987) increased the sensitivity of their assay to the point where Z-DNA formation could be detected.

Like Z-DNA, cruciform formation is thermodynamically favourable only in negatively supercoiled DNA. Consequently, the detection of cruciforms *in vivo* using M*Eco*RI would similarly be hampered by the overmethylation of newly-replicated, relaxed linear DNA. In order to pre-empt such difficulties, a temperature-sensitive M*Eco*RI^{ts} enzyme was considered as a probe of cruciform extrusion *in vivo*. The activity of this methylase at different temperatures was characterised using the plating of modified λ phage on an *Eco*RI restricting host. These results are presented below, followed by a discussion of the merits of M*Eco*RI^{ts} as a probe of cruciform extrusion, in particular when used in conjunction with a λ vector for long DNA palindromes.

Results

Growth of DL559 (pLV59) at Restrictive Temperature

The gene encoding the M*Eco*RI^{ts} enzyme is carried by the plasmid pLV59 (O'Connor & Humphreys, 1982), which was maintained in the *E. coli* strain DL559 by chloramphenicol selection. pLV59 also carries the gene for the *Eco*RI restriction endonuclease (R*Eco*RI), which is expressed independently of the temperature-sensitive methylase. The R*Eco*RI may be inactivated by insertion of a DNA fragment at unique *Hind*III, *Bgl*II or *Pst*I sites located in the gene. This allows the hybrid plasmid to survive at 37°C or 42°C.

Although no such DNA manipulations had been previously performed on the pLV59 used here, the *E. coli* DL559 carrying this plasmid was able to grow at the restrictive temperatures of 37°C and 42°C. The two simplest explanations for this phenomenon are:

1. the R*Eco*RI had been inactivated,
2. or the mutant M*Eco*RI had reverted to wild-type and was no longer temperature-sensitive.

These possibilities were investigated by the growth of λ phage in DL559, and the subsequent restriction of plating on R*Eco*RI hosts.

Restriction of λ NEM43 Plating on DL558 (R*Eco*RI$^+$)

Bacteriophage λ strain NEM43 (*b538 cIam*) was used to assay the methylase activity in DL559. λ strain NEM378 was used as a control: apart from lacking all *Eco*RI sites, it is isogenic to NEM43. Unlike NEM43, NEM378 will plate on an R*Eco*RI host. This was confirmed using *E. coli* strain DL558, which carries the plasmids pLV56 and NTP14. pLV56 (O'Connor & Humphreys, 1982) contains the wild-type M*Eco*RI gene, and NTP14 (Smith *et al.*, 1976) contains both the wild-type M*Eco*RI and wild-type R*Eco*RI genes. The plasmid-less parental *E. coli* strain HB101 (Boyer & Roulland-Dussoix, 1969) was used as a control. The results presented in Table 3.1 show that unmodified NEM43 suffers a 225-fold restriction in plating on DL558, whereas NEM378 plates with nearly equal efficiency on DL558 and HB101.

Plaques formed by growth of NEM43 on a DL558 host were picked; phage isolated from these plaques had been modified by growth in an M*Eco*RI host, and should consequently plate with increased efficiency on an R*Eco*RI strain. This was

confirmed by growth of the modified NEM43 in DL558, and as can be seen in Table 3.1 the modification by M*Eco*RI restores the plating efficiency of NEM43 on an R*Eco*RI host to the same level as the NEM378 control.

Modification of λ NEM43 by M*Eco*RI^{ts} Activity in DL559

The efficiency of NEM43 modification by the temperature-sensitive methylase in DL559 was similarly analysed. Unmodified NEM43 and NEM378 were grown in DL559 at 30°C, 37°C and 42°C. Phage were isolated from plaques formed on the host DL559 lawn, and assayed for modification by plating with the restricting DL558 strain. HB101 was used as a control host strain. The results are presented in Table 3.2 and show that growth in DL559 at 30°C alleviates the restriction of plating on DL558. A reduction in the plating efficiency of NEM43 phage that had been grown in DL559 at 37°C and 42°C can also be seen in DL558. The plasmid pLV59 in DL559 therefore contains an inactive R*Eco*RI gene, rather than a wild-type M*Eco*RI reversion; the methylase activity is clearly temperature-sensitive.

The efficiency of plating of M*Eco*RI^{ts}-modified NEM43 was compared on DL559 and HB101, using the efficiency of plating of NEM378 to correct for any differences that are not a consequence of *Eco*RI restriction or modification. These calculations are both presented and explained in greater detail in Table 3.3 and Fig. 3.1, and express the comparative efficiency of plating as a "coefficient of modification". The control value in Table 3.3 was calculated using unmodified NEM43 and represents the proportion of phage that escape restriction by R*Eco*RI. As the coefficient of modification is greater for NEM43 phage that have been grown in DL559 at 42°C than for unmodified NEM43 phage, it may be concluded that the temperature-sensitive methylase activity is not completely abolished at 42°C. Furthermore, λ phage grown in DL559 at 37°C still show substantial modification, indicating that the M*Eco*RI^{ts} in pLV59 retains significant activity at what is normally considered a restrictive temperature.

Table 3.1

Restriction of λ NEM43 Plating Efficiency on DL558

Table showing the plating behaviour of NEM378 and NEM43 on DL558 and HB101 host strains. The data are given as plaque forming units (PFU) per ml, and represent the mean of results of two separate plating experiments. DL558 contains plasmids encoding the *Eco*RI restriction/modification system, whereas HB101 is the isogenic plasmid-less strain. Unmodified NEM43 shows a 225-fold reduction of plating efficiency on DL558 compared to HB101, whereas NEM378 show no significant difference in plating behaviour; NEM378 has no *Eco*RI sites and is consequently unaffected by *Eco*RI restriction/modification systems. Modified NEM43 was isolated from plaques formed on a lawn of DL558 cells, and shows an increase in plating efficiency on DL558 over unmodified NEM43.

Table 3.2

Modification of λ NEM43 by the M*Eco*RI^{ts} Activity in DL559

Table showing the plating behaviour of M*Eco*RI^{ts}-modified NEM43 and NEM378 on DL558 and HB101 host strains. The data are given as PFU ml^{-1}, and represent the mean of results of two separate plating experiments. NEM43 (and the NEM378 control) were modified by the plasmid-encoded M*Eco*RI^{ts} activity in DL559 at the different temperatures shown in the Table. The extent of modification was subsequently assayed by the plating efficiency on DL558 and HB101 hosts. A decrease in plating efficiency of NEM43 on DL558 relative to HB101 represents the restriction of unmodified phage; this is most apparent for NEM43 phage that have been isolated from DL559 grown at 42°C. Such a reduction in plating efficiency is not seen for NEM378, which is unaffected by *Eco*RI restriction/modification.

Table 3.1

		E. coli Strain	
		DL558 (*Eco*RI restriction$^+$/modification$^+$)	HB101
λ Strain	NEM378 No *Eco*RI Sites	2.1×10^9	3.7×10^9
	NEM43 Unmodified	8.0×10^4	1.8×10^7
	NEM43 M*Eco*RI Modified	2.8×10^5	5.0×10^5

Table 3.2

		E. coli Strain					
		DL558 (*Eco*RI restriction$^+$/modification$^+$)			HB101		
Temperature of Modification (in DL559)		30°C	37°C	42°C	30°C	37°C	42°C
λ Strain	NEM43 M*Eco*RI^{ts} Modified	2.5×10^5	1.0×10^5	1.3×10^3	2.9×10^5	2.9×10^5	7.0×10^4
	NEM378 No *Eco*RI Sites	2.4×10^5	3.0×10^5	1.6×10^4	3.0×10^5	3.2×10^5	2.3×10^4

Table 3.3

Modification of λ NEM43 by M*Eco*RI^{ts} at Different Temperatures

Table showing the extent of modification of NEM43 by the M*Eco*RI^{ts} activity in DL559 at different temperatures. The values were calculated from the data in Table 3.2 using the equation:

$$\text{Coefficient of Modification} = \frac{\left(\dfrac{\text{Titre of NEM43 } [t^{\circ}\text{C}] \text{ on DL558}}{\text{Titre of NEM43 } [t^{\circ}\text{C}] \text{ on HB101}}\right)}{\left(\dfrac{\text{Titre of NEM378 } [t^{\circ}\text{C}] \text{ on DL558}}{\text{Titre of NEM378 } [t^{\circ}\text{C}] \text{ on HB101}}\right)} \qquad (8)$$

The upper part of the equation calculates the restriction of NEM43 plating on DL558 that is due to incomplete modification by the M*Eco*RI^{ts} activity in DL559 at the temperature t. The lower half of the equation uses the plating behaviour of the control λ phage NEM378 to correct for any differences in plating efficiency on DL558 and HB101 that are not a consequence of *Eco*RI restriction/modification. The control value was calculated using the data in Table 3.1 corresponding to the restriction of unmodified NEM43 and NEM378 plating on a DL558 host. It therefore represents the basal coefficient of modification.

Figure 3.1

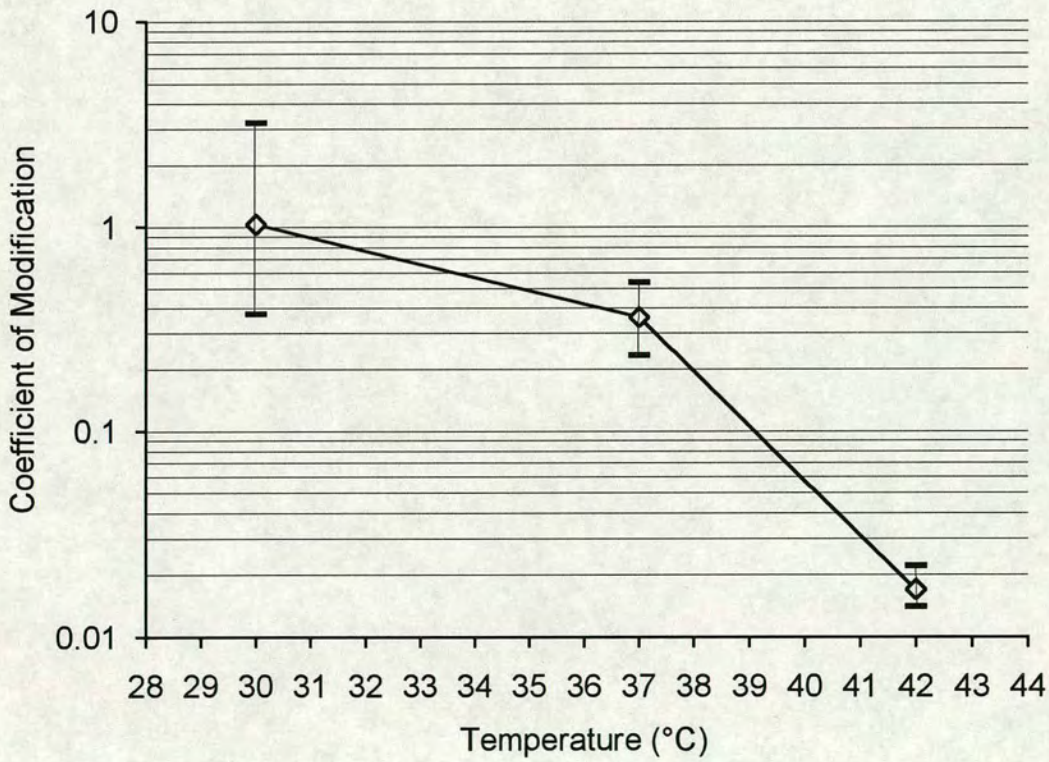
Coefficient of Modification by M*Eco*RI^{ts} vs. Temperature

Graph of the coefficient of modification of NEM43 by the M*Eco*RI^{ts} activity in DL559 vs. temperature of modification. This graph represents the values shown in Table 3.3, after correction for the basal coefficient of modification: the coefficient of modification for unmodified NEM43 was subtracted from the coefficients of modification of NEM43 by M*Eco*RI^{ts}. Error bars are also shown, representing the range of coefficients of modification at various temperatures as calculated from individual plating experiments.

Table 3.3

Temperature (°C)	Coefficient of Modification
30	1.042
37	0.368
42	2.46×10^{-2}
Control	7.68×10^{-3}

Figure 3.1



Conclusion

The results presented above show that the *Eco*RI methylase activity encoded by the pLV59 plasmid in DL559 is clearly temperature-sensitive, and may therefore be suitable as a probe of DNA secondary structure *in vivo*. However, the data in Table 3.3 and Figure 3.1 demonstrates that the M*Eco*RI^{ts} retains significant activity at 37°C, although this is normally considered a restrictive temperature. This implies that the cells should ideally be grown at 42°C to prevent methylation of newly-replicated linear relaxed DNA, before undergoing a temperature change to 30°C to activate the methylase probe.

Such a dramatic shift in temperature may not be compatible with the other components of the assay for cruciform extrusion presented here. The palindromic sequence used here is carried by a λ phage derivative, and the growth of this vector would be compromised by the high restrictive temperature of 42°C; rapid temperature changes from 42°C to 30°C would further reduce the viability of the phage vector. Furthermore, a high temperature could significantly alter the kinetics of cruciform extrusion, which would undermine the non-invasive credentials of the assay. While the use of temperature-sensitive methylase probes in conjunction with plasmids appears to cause no problems, it is highly probable that such a method could not be transferred to a phage-based system without serious complications. M*Eco*RI^{ts} was therefore not chosen as a probe of cruciform extrusion *in vivo*; Dam methylase was used instead, for the reasons given in the next chapter.

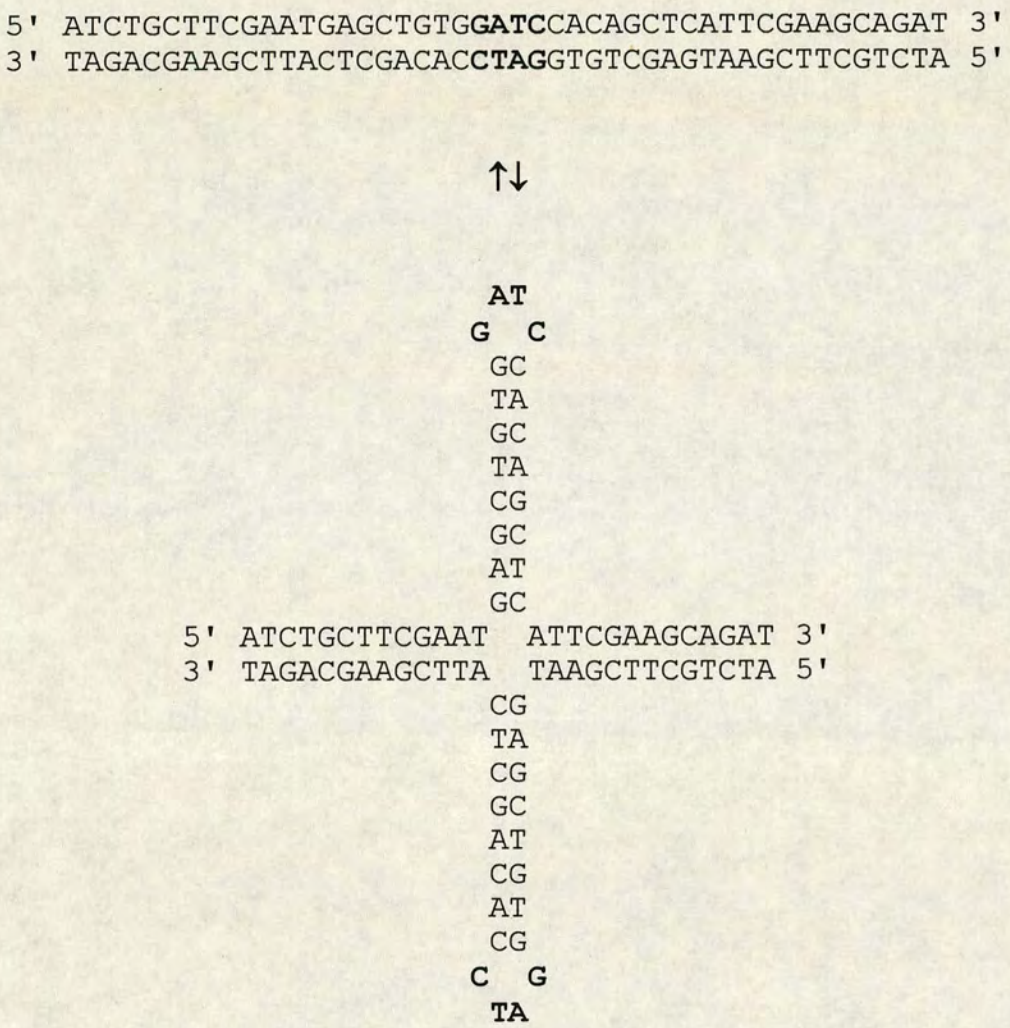
In summary, this chapter shows that the M*Eco*RI^{ts} activity encoded by pLV59 retains significant activity at 37°C, and may therefore be unsuitable as an inducible probe of cruciform extrusion *in vivo*.

CHAPTER 4

ANALYSIS OF DAM METHYLATION AT A PALINDROME CENTRE *IN VIVO*

Figure 4.1

Partial extrusion of a cruciform structure from a palindromic DNA sequence with a central GATC target site for Dam methylase.



Introduction

Dam Methylase as a Probe of Cruciform Extrusion

Dam methylase is an endogenous methylase of *E. coli*, which modifies GATC target sites at adenine residues (reviewed by Barras & Marinus, 1989). This methylation is essential for methyl-directed DNA mismatch correction (reviewed by Modrich, 1989). Dam methylase has been used as an *in vivo* probe of triplex H-DNA formation in *E. coli* (Parniewski *et al.*, 1990), and to investigate DNA-protein interactions in *E. coli* (Wang & Church, 1992) and chromatin structure in *Saccharomyces cerevisiae* (Singh & Klar, 1992). Dam methylase was chosen in preference to *Eco*RI methylase as a probe of cruciform extrusion *in vivo* for three principal reasons:

1. Dam methylase is endogenous to *E. coli*,
2. both methylated and unmethylated Dam targets can be independently quantified using the methylation-sensitive restriction endonucleases *Dpn*I and *Mbo*I respectively (*Eco*RI methylation can only be quantified with the *Eco*RI restriction endonuclease),
3. and the Dam methylase target site is only 4 bp in size (as opposed to the *Eco*RI target, which is 6 bp), allowing greater flexibility in changes to the palindrome central sequence.

Inhibition of DNA Methylation by Cruciform Loops

The assay used here relies on the observation that specific modification enzymes are unable to modify their target sites when these are in regions of unusual DNA secondary structure. In this case, the unusual DNA structure is the single-stranded loops that are formed by an extruded cruciform; these cruciform loops contain the central bases of the palindromic DNA sequence. If a palindrome with a central GATC target site for Dam methylase forms a cruciform *in vivo*, then this target will be located in the single-stranded loops and will consequently not be modified (Fig. 4.1). The ability of cruciform loops to inhibit DNA methylases has been shown *in vitro* (Murchie & Lilley, 1989). After DNA purification, the degree of methylation at the GATC sites is examined by cleavage with methylation-sensitive restriction endonucleases such as *Dpn*I which cuts only completely methylated GATC, and *Mbo*I which cuts only completely unmethylated GATC.

Bacteriophage λ as a Vector for Long DNA Palindromes

As described in the Introduction, long DNA palindromes lead to the inviability to their carrier replicon in wild-type *E. coli*. While the use of *recBC sbcB(C)* mutants allows the plating of λ phage containing long palindromes (Leach & Stahl, 1983), such hosts do not permit the stable propagation of plasmids (Cohen & Clark, 1986). Bacteriophage λ has therefore traditionally been the vector of choice for long palindromes. The finding that mutation of the *sbcC* gene is primarily responsible for the alleviation of palindrome-mediated inviability and that plasmids are stably maintained in *sbcC* hosts (Chalker *et al.*, 1988) has enabled the cloning of palindromic DNA sequences in plasmid vectors. However, even in an *sbcC* host, palindromes appear to be more stable in λ vectors than in plasmids (A. Chalker, Ph.D. Thesis, 1990). Consequently, the palindromes used in the assay described here are carried by a λ vector.

The use of a λ vector has two additional advantages over plasmids:

1. λ DNA is large molecule (48.5 kb for wild-type λ), providing a high number of GATC sequences outside the palindrome that act as control target sites,
2. and λ replicates very rapidly, with a generation time comparable to the half-time of methylation by Dam methylase; this prevents overmethylation and therefore increases the sensitivity of the assay.

Both of these points are discussed in greater detail later in this chapter and in subsequent chapters.

Results

Insertion of GATC Site at the Centre of a Palindrome

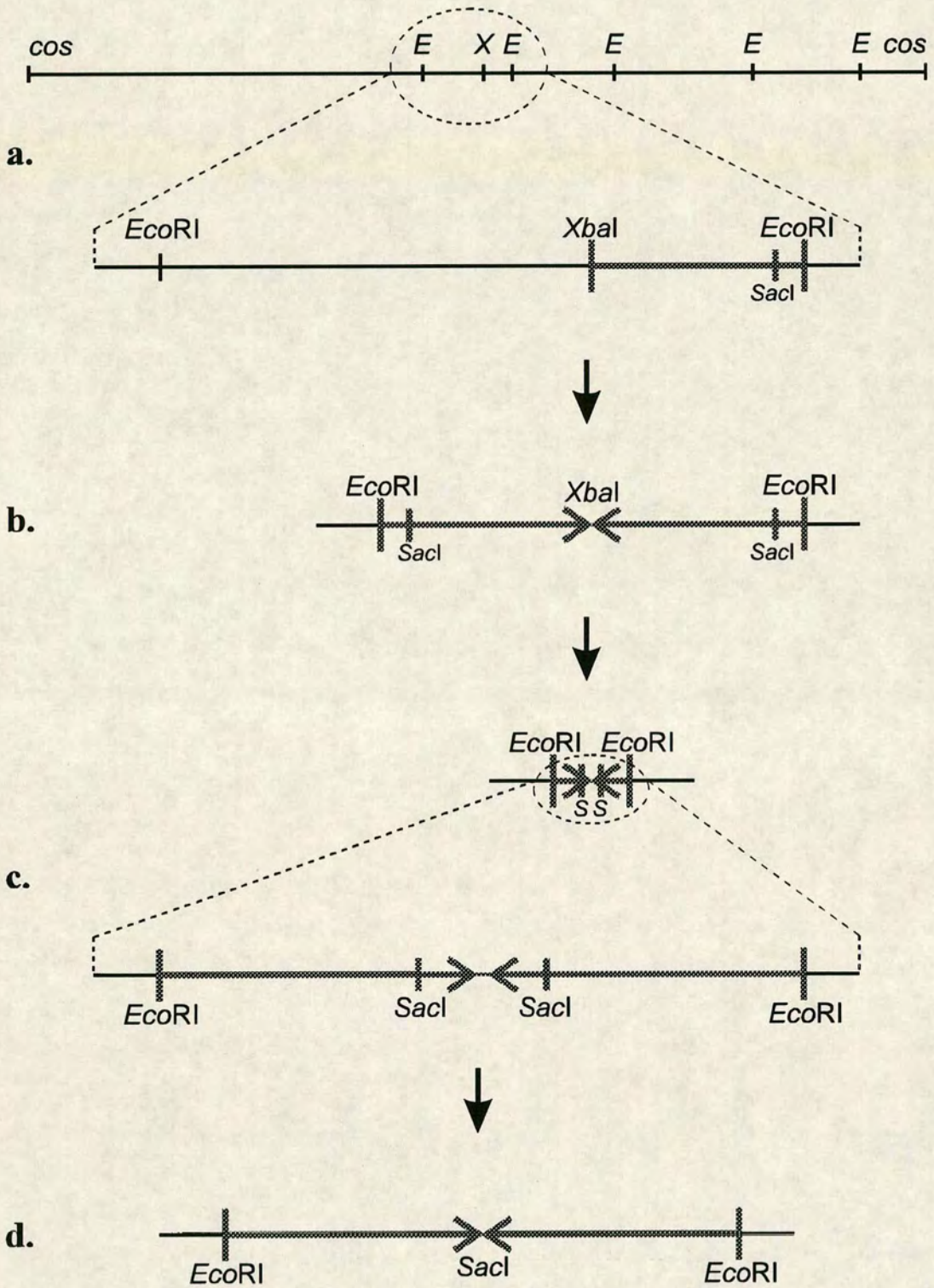
In order to generate a long perfect DNA palindrome with a central GATC site, an oligonucleotide with such a target for Dam methylase was inserted at the central *Sac*I restriction endonuclease site of the 462 bp palindrome in λ phage DRL133.

Palindromic DNA Sequence in DRL133

The construction of the 462 bp palindromic DNA sequence in DRL133 has been described previously (Chalker *et al.*, 1993). It is briefly summarised below with reference to publications reporting the relevant construction steps, and is illustrated schematically in Fig. 4.2.

Figure 4.2

Scheme illustrating the construction of the 571 bp palindrome in λ DRL116 and the 462 bp palindrome in λ DRL133 (see text for details).



Initially, a perfect palindrome of 3.2 kb was constructed between *Eco*RI sites 1 and 2 of phage λ using the 1.6 kb λ *Xba*I-*Eco*RI fragment, thus replacing the non-essential *Eco*RI B fragment (Leach & Stahl, 1983). This is illustrated in steps a. and b. of Fig. 4.2, where *X* represents the *Xba*I site, *E* the *Eco*RI sites, and *cos* the cohesive ends of λ. The 1.6 kb *Xba*I-*Eco*RI fragment is shown in grey ( ). A 571 bp palindrome was constructed by an *in vivo* deletion from this 3.2 kb palindrome (Leach & Lindsey, 1986) as shown in steps b. to c. of Fig. 4.2; the duplicated *Sac*I site is designated by *S*. This 571 bp palindrome is present in λ DRL116 (Δ*B pal*571 *spi*6 *cI*857) (Chalker *et al.*, 1988) and has been analysed by DNA sequencing (J. Lindsey, Ph.D. Thesis, 1987): it contains a central asymmetry of 15 bp, and the duplicated *Sac*I sites are separated by 109 bp. The 462 bp palindrome in DRL133 was constructed by removal of this 109 bp *Sac*I-*Sac*I fragment from the palindrome of DRL116, generating a perfect palindrome with a central *Sac*I site that is unique in DRL133 (Chalker *et al.*, 1993). The construction of this palindrome is illustrated by steps c. to d. of Fig. 4.2. The complete sequence of the 462 bp palindrome in DRL133 is shown in the Appendix to this thesis.

Palindromic Oligonucleotide with Central GATC Site

This unique *Sac*I site facilitates the insertion of oligonucleotides at the palindrome centre. In the case of the work described in this chapter, the oligonucleotide contains a central GATC target site. It is shown below:

```
                Dam
5'      GTGGATCCACAGCT 3'
3' TCGACACCTAGGTG      5'
   SacI    BamHI    SacI
```

The GATC Dam methylase site additionally forms part of a GGATCC target sequence for *Bam*HI restriction endonuclease. The presence of a new *Bam*HI site in λ facilitates the identification of recombinant palindromes. The oligonucleotide has 3' terminal linker sequences (*Sac*I) that are complementary to the four-base overhang produced by *Sac*I cleavage. The presence of a C residue 5' to this linker inactivates the GAGCTC *Sac*I target site upon insertion, thereby facilitating the isolation and identification of desired clones.

Cloning of Oligonucleotide with GATC Site in Palindrome Centre

The oligonucleotide with a central GATC site was heated to 100°C and allowed to cool to room temperature slowly, thereby favouring the formation of

double-stranded DNA over self-annealed hairpins. It was then inserted by ligation at the centre of the palindrome in DRL133, which had previously been cleaved to completion with *Sac*I. The oligonucleotide was not phosphorylated, thus preventing the insertion of multimers, and was ligated with the DRL133 DNA at a 3:1 (oligonucleotide:λ) molar ratio of cohesive *Sac*I termini. After ligation, the DNA ligase was inactivated at 70°C and the reaction was allowed to cool slowly to room temperature as before. The DNA was cleaved to completion for a second time with *Sac*I in order to select against phage without an oligonucleotide insertion at the palindrome centre. The DNA was precipitated with ethanol, resuspended in 5μl of TE buffer, and *in vitro* packaged to produce mature λ particles.

Selection of λ Phage with Oligonucleotide Insertion in Palindrome

Although the second round of cleavage with *Sac*I should select against all phage that failed to clone the oligonucleotide at the palindrome centre, it is ineffective against phage that have undergone a deletion of part or all of the palindrome. As palindromic sequences are notoriously unstable, care must be taken to eliminate such false recombinants. This is achieved by selecting recombinant phage which plate normally on an *sbcC* mutant but fail to plate on a wild-type host; these phage will have retained the long DNA palindrome.

The *in vitro*-packaged phage were amplified by one round of plating on a JC9387 (*recBC sbcBC*) host. A *recBC* mutant host was used to permit rolling circle replication by a *spi*⁻ (*red gam*) χ° λ phage such as the DRL133 derivative described here; the significance of χ to λ replication, and a possible influence on cruciform extrusion is presented in the next chapter. Approximately 50 plaques were picked into phage buffer and spotted onto bacterial lawns of 594 (*rec*$^+$) and JC9387 (*recBC sbcBC*). Phage growing on JC9387 but not on 594 were selected, and purified by two further rounds of growth on JC9387. Single plaques were picked for the preparation of phage stocks by plate lysates, again using JC9387 as a host strain.

λ DNA was prepared from the plate lysates and cut with *Sac*I to identify false recombinant phage which had escaped cleavage with *Sac*I prior to *in vitro* packaging. The phage which failed to cut in this third round of *Sac*I digestion were cleaved to completion with *Bam*HI to identify the presence of the oligonucleotide insertion. The phage were additionally cleaved with *Stu*I to generate DNA fragments that are easily discernible by agarose gel electrophoresis. The introduction of a novel *Bam*HI site in an 11.1 kb *Stu*I-*Bam*HI fragment, generating 9 kb and 2.1 kb fragments, was taken to indicate that the cloning had been successful. The phage was designated DRL151.

Analysis of *in Vivo* Dam Methylation

In Vivo Methylation by Growth of λ with Palindrome in *E. coli* Host

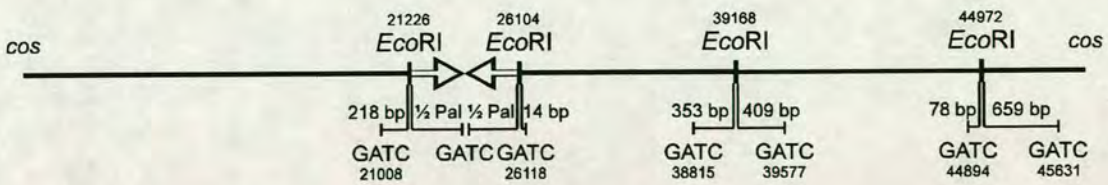
DRL151, the λ phage carrying the palindrome with the GATC centre, was used to infect JC9387 (*recBC sbcBC dam*$^+$) at a multiplicity of infection (MOI) of 0.1, and grown by liquid lysis. After complete cell lysis, phage particles were purified and λ DNA extracted using the large scale method. The numerous rounds of growth of DRL151 in the JC9387 host ensure that the level of Dam methylation at GATC targets reflects the steady state concentrations of methylation-sensitive and resistant structures at those sites.

Analysis of Methylation at GATC Sites in DRL151

In order to assay the extent of Dam methylation at the palindrome centre, the λ DNA can be cleaved with two complementary methylation-sensitive restriction endonucleases: *Dpn*I (Lacks & Greenberg, 1975; Vovis & Lacks, 1977), which cuts only fully methylated GATC, and *Mbo*I (Gelinas *et al.*, 1977), which cuts only fully unmethylated GATC. However, there are 116 GATC targets in (wild-type) λ, and many more DNA fragments than this will be formed by what is essentially a methylation-dependent partial restriction digest. In order to simplify the analysis of Dam methylation, the λ DNA was initially cut to completion with *Eco*RI and radiolabelled at these *Eco*RI sites using [α^{32}P]dATP and Klenow enzyme. If quantification is confined to DNA fragments with radiolabelled *Eco*RI termini, then the measurement of Dam methylation is greatly simplified. Cleavage by *Dpn*I or *Mbo*I at GATC sites that are immediately adjacent to the radiolabelled *Eco*RI sites produces a discrete number of DNA fragments of predictable sizes, which can be quantified by autoradiography and densitometry. The intensity of these fragments will be directly proportional to the efficiency of *Dpn*I or *Mbo*I cleavage at that site. The position of these GATC targets and the sizes of predicted fragments are illustrated in Fig. 4.3.

Figure 4.3

Map of λ DRL133 (not to scale) showing the GATC sites assayed for methylation (numbers indicate positions in the wild-type λ sequence (Sanger *et al.*, 1982)).



The GATC sites lying outside the palindrome represent control targets for Dam methylase, as they are not expected to adopt a methylation-resistant DNA structure *in vivo*. The methylation of the palindrome centre can be compared to these other GATC sites, thereby ensuring that the experiment is internally controlled. *Eco*RI sites were used for radiolabelling as DRL133 contains only four of these targets, two of which flank the palindromic sequence. The fragments resulting from cleavage at *Eco*RI sites and at adjacent GATC sites are nearly all discernible by polyacrylamide gel electrophoresis; the size of the half palindrome (½ Pal) is 234 bp. The presence of the oligonucleotide insertion at the palindrome centre can be verified by cleavage with *Bam*HI in conjunction with the ^{32}P-labelling of *Eco*RI termini, again producing a 234 bp half palindrome fragment.

After radiolabelling, the DNA was extracted with phenol-chloroform and chloroform, precipitated with ethanol and resuspended in TE buffer. The buffer was appropriately adjusted, and the DNA was cleaved with either *Bam*HI, *Dpn*I, or *Mbo*I, or was left uncut. The optimal ratio of *Eco*RI and *Bam*HI restriction endonuclease to λ DNA had been determined experimentally to be approximately 5 units (U) enzyme per μg λ DNA. Owing to the methylation sensitivity of *Dpn*I and *Mbo*I, such an optimisation is not possible; these endonucleases were therefore used at a similar concentration of 5 U μg^{-1} λ DNA. After the restriction digest, the DNA was extracted with phenol-chloroform and chloroform, precipitated with ethanol and resuspended in formamide-EDTA gel-loading buffer. Samples were electrophoresed on a 6% denaturing polyacrylamide gel; denaturing conditions, including a uniform gel temperature of 50 – 55°C, were found to be necessary to prevent hairpin formation by the full-length palindrome. After electrophoresis, the gel was dried and autoradiographed.

Quantification of Methylation at GATC Sites

The autoradiograph resulting from such an assay for Dam methylation of DRL151 is shown in Fig. 4.4. The bands labelled in Fig. 4.4 were quantified by densitometry and the values are shown in Table 4.1. Also shown in Table 4.1 are the results of calculations determining the degree of Dam methylation at various GATC sites *in vivo*, and the relative efficiency of cleavage by *Dpn*I and *Mbo*I at these sites. The latter values are also presented graphically in Figs. 4.5 and 4.6. While the values in Table 4.1 only describe one experiment, the analysis of Dam methylation of λ DRL151 DNA has been repeated a number of times with very similar results (data not shown). The relevant conclusions are discussed below.

Figure 4.4

Autoradiograph of Restriction Digests at GATC Sites in λ DRL151

Autoradiograph showing restriction digests of λ DRL151 DNA determining the level of Dam methylation at various GATC sites *in vivo*. The λ phage had been grown in JC9387 and DNA purified as described in the text. The DNA was cleaved with *Eco*RI and radiolabelled with [α^{32}P]dATP as described in the text, before cleavage with *Bam*HI, *Dpn*I, *Mbo*I, *Dpn*I and *Mbo*I, or *Sau*3AI. The appropriate enzymes are indicated above each lane. The full-length and half-length palindrome, and the control fragments described in the text (and Fig. 4.3) are shown by the labels on the right of the autoradiograph. The two lanes on the left of the autoradiograph show DNA fragment size markers generated by a restriction digest of plasmid pBR322 with *Hpa*II, followed by radiolabelling using [α^{32}P]dCTP and Klenow enzyme. The sizes of the DNA fragments are indicated by the labels on the left of the autoradiograph.

Digestion with only *Eco*RI generates the full length palindrome, while additional cleavage by *Bam*HI at the central GGATCC site generates the half-length palindrome. Digestion with *Dpn*I indicates the level of fully methylated DNA, and *Mbo*I the level of fully unmethylated DNA. Cleavage with both *Dpn*I and *Mbo*I was carried out to ascertain the level of hemimethylated DNA which will be cut by neither enzyme; the results were not used in later analysis. *Sau*3AI reportedly cleaves DNA to completion at GATC sites regardless of Dam methylation, and was used here to indicate the size of the control DNA fragments. However, despite optimisation of the incubation conditions and the enzyme to DNA ratio, it was found to be impossible to eliminate the appearance of faint bands (presumably resulting from partial digestion) in the restriction digest. The use of *Sau*3AI was consequently discontinued.

From left to right, the lanes are:

1. pBR322 *Hpa*II DNA fragment size marker
2. pBR322 *Hpa*II DNA fragment size marker (1/3 dilution)
3. λ DRL151 *Eco*RI
4. λ DRL151 *Eco*RI *Bam*HI
5. λ DRL151 *Eco*RI *Dpn*I
6. λ DRL151 *Eco*RI *Mbo*I
7. λ DRL151 *Eco*RI *Dpn*I *Mbo*I
8. λ DRL151 *Eco*RI *Sau*3AI

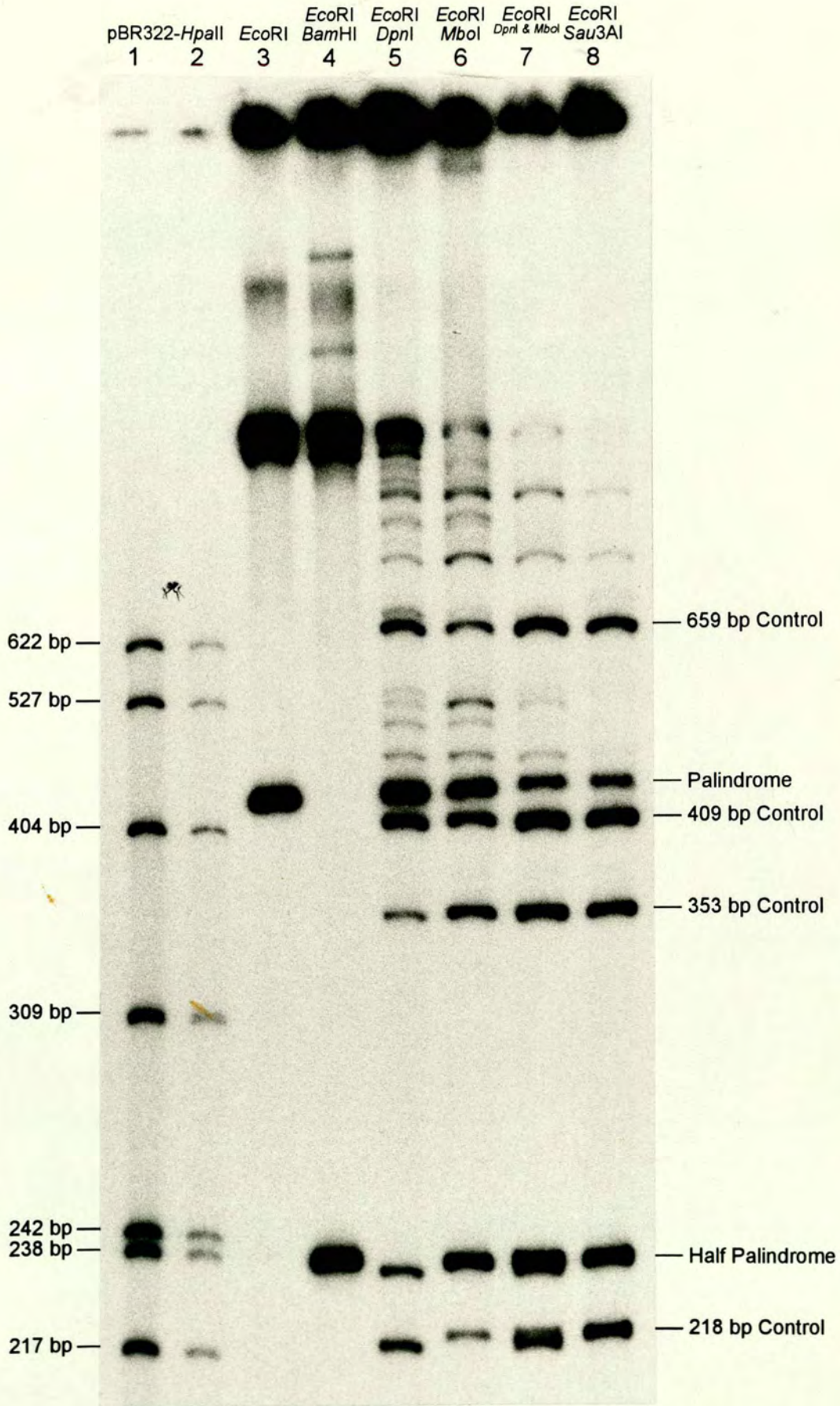

pBR322-HpaII
EcoRI
EcoRI BamHI
EcoRI DpnI
EcoRI MboI
EcoRI DpnI & MboI
EcoRI Sau3AI
1
2
3
4
5
6
7
8
622 bp
527 bp
404 bp
309 bp
242 bp
238 bp
217 bp
659 bp Control
Palindrome
409 bp Control
353 bp Control
Half Palindrome
218 bp Control

Table 4.1 Methylation of GATC Sites in DRL151 *in Vivo*

Table showing data derived from the autoradiograph in Fig 4.4. The bands shown in Fig. 4.4 were quantified by densitometry and their intensities are displayed here in rows designated "Data". The quantified values (Uncorrected Data) for the full-length and half-length palindrome were halved (Corrected Data), as the palindrome is radiolabelled at both ends whereas the control fragments are radiolabelled at only one end. The % cleavage at the palindrome centre by *Dpn*I or *Mbo*I was calculated by:

$$\%\text{Cleavage}_{\text{Pal}} = \frac{\text{½Pal}}{(\text{½Pal} + \text{Pal})} \times 100 \qquad (9)$$

The % methylation at the palindrome centre was determined from the levels of methylated DNA as measured by *Dpn*I and *Mbo*I cleavage. The values from both enzymes were averaged to take into account hemimethylated DNA which is cleaved by neither enzyme, and were calculated by:

$$\%\text{Methylation}_{\text{Pal}} = \frac{\left(\frac{\text{½Pal}_{\text{Dpn}}}{\text{½Pal}_{\text{Dpn}} + \text{Pal}_{\text{Dpn}}}\right) + \left(1 - \frac{\text{½Pal}_{\text{Mbo}}}{\text{½Pal}_{\text{Mbo}} + \text{Pal}_{\text{Mbo}}}\right)}{2} \times 100 \qquad (10)$$

The average % cleavage at control sites cannot be determined directly, as the fraction of uncut control DNA cannot be measured. Instead, the intensity of control bands is compared to the sum of the full-length and half-length palindrome band intensities, which should be labelled at each of its *Eco*RI termini at the same efficiency as the control fragments (the intensity of the palindrome fragments is halved as described above). Using the sum of the palindrome bands as a standard,

$$\%\text{Cleavage}_{\text{Control}} = \frac{\text{AverageControl}}{(\text{½Pal} + \text{Pal})/2} \times 100 \qquad (11)$$

The average % methylation of the control sites is similarly calculated using the sum of the palindrome bands as a standard. The *Dpn*I and *Mbo*I values are averaged, as for the GATC site at the palindrome centre.

$$\%\text{Methylation}_{\text{Control}} = \frac{\left(\frac{\text{AverageControl}_{\text{Dpn}}}{(\text{½Pal}_{\text{Dpn}} + \text{Pal}_{\text{Dpn}})/2}\right) + \left(1 - \frac{\text{AverageControl}_{\text{Mbo}}}{(\text{½Pal}_{\text{Mbo}} + \text{Pal}_{\text{Mbo}})/2}\right)}{2} \times 100 \qquad (12)$$

The bottom portion of the table shows the relative efficiency of cleavage by *Dpn*I and *Mbo*I at the palindrome and control GATC sites. The values were calculated relative to the average intensity of the control GATC fragments, which is shown with the control data as "Average Control".

Table 4.1

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$	*Mbo*I (GATC)
Uncorrected Data	Palindrome	25603	17596
	½ Palindrome	6690	18568
Corrected Data	Palindrome	12802	8798
	½ Palindrome	3345	9284
% Cleavage at Palindrome Centre		20.7	51.3
% Methylation of Palindrome Centre		34.7	
Control Data	659 bp Control	9303	5307
	409 bp Control	10746	8350
	353 bp Control	3835	8464
	218 bp Control	7389	3954
	Average Control	7818	6519
Average % Cleavage at Control Sites		48.4	36.1
Average % Methylation of Control Sites		56.2	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	42.8	142.4
	669 bp Control	119	81.4
	409 bp Control	137.4	128.1
	353 bp Control	49.1	129.9
	218 bp Control	94.5	60.7
	Average Control	100	100

Figure 4.5

Relative Cleavage by *Dpn*I at GATC Sites in DRL151

Graph showing the relative efficiency of cleavage by *Dpn*I at GATC sites in λ DRL151 following Dam methylation *in vivo* in JC9387. The values plotted are those shown on the left-hand side of the bottom portion of Table 4.1. The intensities of the bands shown in Fig. 4.4 are presented relative to the average intensity of the control bands, although the value for the half-length palindrome band was halved to correct for radiolabelling at two *Eco*RI sites. Error bars are also shown, representing the range of relative cleavage by *Dpn*I at GATC sites as calculated from individual restriction digests. Figure 4.5 represents graphically the presence of Dam methylation at various GATC sites in DRL151.

Figure 4.6

Relative Cleavage by *Mbo*I at GATC Sites in DRL151

Graph showing the relative efficiency of cleavage by *Mbo*I at GATC sites in λ DRL151 following Dam methylation *in vivo* in JC9387. The values plotted are those shown on the right-hand side of the bottom portion of Table 4.1. The data are presented as in Fig. 4.5. Error bars are also shown, representing the range of relative cleavage by *Mbo*I at GATC sites as calculated from individual restriction digests. Figure 4.6 represents graphically the absence of Dam methylation at various GATC sites in DRL151.

Figure 4.5

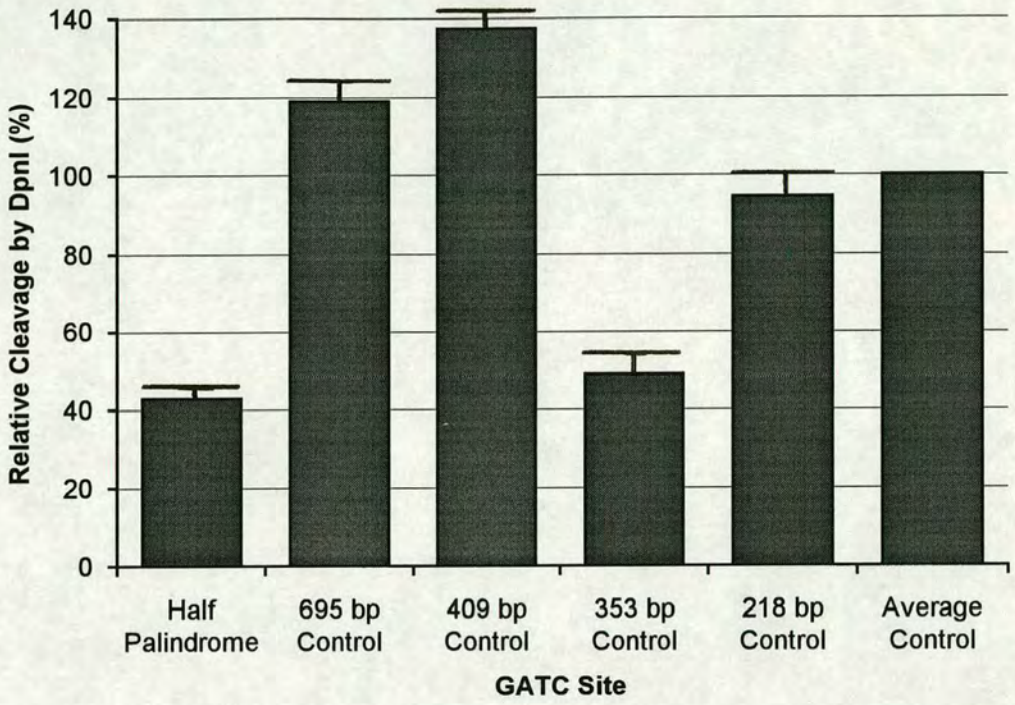
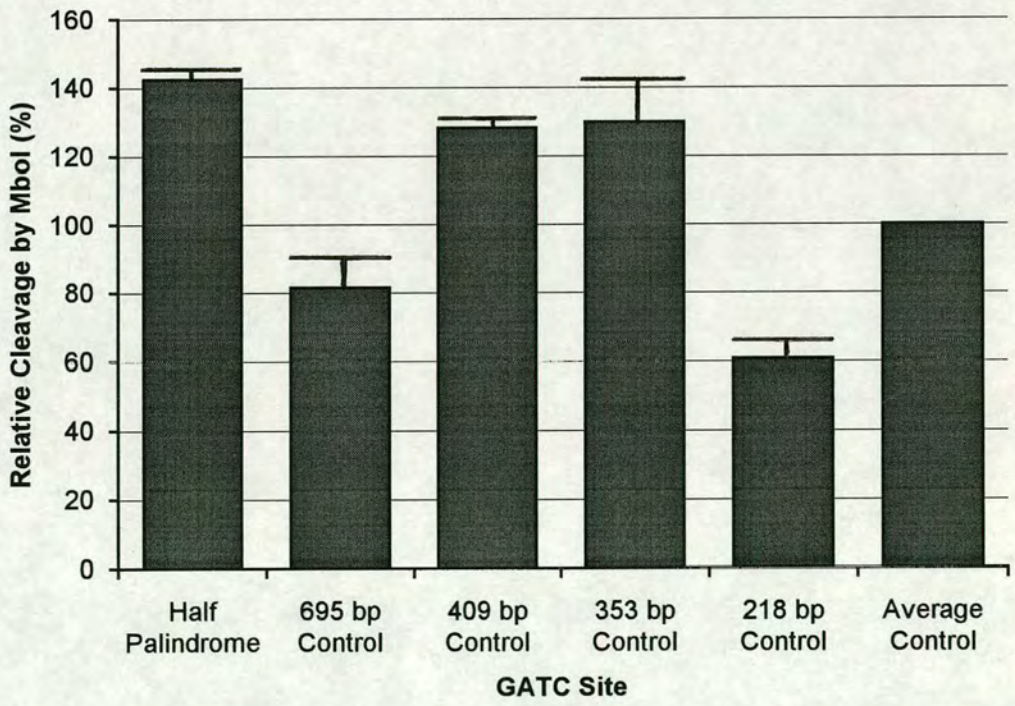


Figure 4.6



Conclusion

The results presented in Table 4.1 show that the palindrome centre is undermethylated compared to the control GATC sites. The palindrome GATC target is methylated by 34.7%, whereas the control GATC sites are methylated by an average of 56.2%. These values were determined by a calculation which takes into account cleavage by both *Dpn*I and *Mbo*I. Although the oligonucleotide was not methylated at its GATC target prior to insertion into DRL133, it is inconceivable that this undermethylation had persisted through numerous rounds of replication in JC9387. Unlike many other DNA modification enzymes, Dam methylase does not show a significant enhancement of activity on hemimethylated (as opposed to completely unmethylated) DNA (Herman & Modrich, 1982). Furthermore, as explained in the Introduction, self-perpetuating undermethylation of the palindrome centre is prevented by virtue of adenine methylation favouring cruciform extrusion (Murchie & Lilley, 1989). Conversely, the initial undermethylation of the palindrome GATC site will make cruciform extrusion less likely, thereby ensuring that the target site is more accessible to Dam methylase.

In a side-by-side comparison of the efficiency of *Dpn*I and *Mbo*I cleavage at GATC sites (presented in Figs. 4.5 and 4.6), the palindrome centre displays on average less Dam methylation than any of the control GATC sites. However, this undermethylation is not dramatic and may be insignificant given the limitations of this assay. Furthermore, two significant reservations about the results presented here arise as a consequence of the data in Figs. 4.5 and 4.6:

1. the control GATC targets show a high variation in their levels of methylation, which is most probably the result of site-specificity by Dam methylase, and/or *Dpn*I and *Mbo*I,
2. and there is a non-reciprocity in *Dpn*I and *Mbo*I cleavage efficiencies for some control GATC sites, most notably the target giving the 409 bp control fragment.

The first reservation is explored in Chapter 6, which describes an investigation of the effect of sequence context on Dam methylase activity. The second reservation is laid to rest in Chapter 5, which presents evidence that the high intensity of the 409 bp fragment is the result of unwitting quantification of a doublet band.

In summary, while this chapter shows that the GATC site at the centre of a palindrome is poorly methylated *in vivo*, this undermethylation is not substantial and may be insignificant.

CHAPTER 5

EFFECT OF χ^+ ON DAM METHYLATION OF THE PALINDROME CENTRE *IN VIVO*

Introduction

As described in the Introduction to this thesis, a *pal red gam* χ° λ phage such as DRL151 is unable to plate efficiently on a *recBCD*$^+$ host. Bacteriophage lacking a functional *gam* gene product fail to inactivate the host RecBCD enzyme and are therefore unable to produce packageable λ multimers by σ (rolling-circle) replication (Enquist & Skalka, 1973). However, RecBCD enzyme may catalyse the formation of viable λ dimers by homologous recombination; this is stimulated by a χ recombination hotspot (5' GCTGGTGG 3') in the λ sequence (review by Smith, 1988b).

In common with widely-used λ cloning vectors, DRL151 has a *spi6* deletion and is therefore *red gam*; the use of a *red* mutant λ derivative is required to prevent recombination between repeated sequences. Until recently, the inability of such phage to plate on a *recBCD*$^+$ (*sbcC*) host was comparatively unimportant. The viability of λ phages carrying long perfect palindromes is restored by mutations in the *recBC* and *sbcB(C)* genes (Leach & Stahl, 1983), and these mutations additionally permit σ replication by a *red gam* λ phage. However, the finding that mutation of *sbcC* alone is sufficient to overcome the inviability caused by long palindromes (Chalker *et al.*, 1988) has opened up new avenues of research. In order to investigate the behaviour of palindromic DNA in an *sbcC* host, the λ vector carrying the palindrome must be able to replicate efficiently in a *recBCD*$^+$ (*sbcC*) cell. This is most simply achieved by making the λ vector χ^+.

Results

Phage Cross of DRL151 with NEM1239 (χ^+)

DRL151 was crossed with λ NEM1239 (*b*1453 *cI*857 *nin*5 *srI*4° *srI*5° χ^+) to generate a χ^+ derivative of DRL151. If such a cross is promoted by RecBCD-mediated recombination, the χ^+ site is obtained up to 10 times less frequently than its χ° reciprocal (Stahl *et al.*, 1980). This nonreciprocality stems from a bias in packaging of the λ recombinants (Kobayashi *et al.*, 1984). The cross was therefore performed in a JC9387 *recBC sbcBC* host; χ^+ sites are ineffective at stimulating recombination by the RecF pathway that operates in *recBC sbcBC* cells, and should hence be recovered amongst the recombinants with an equal frequency to the χ° reciprocal. To favour the recovery of palindrome-bearing phage amongst the recombinants, JC9387 was infected with DRL151 at a MOI of 10, while NEM1239 was added at a MOI of 2.

λ χ^+ recombinants were identified by plating on an N2364 (*sbcC*) host; χ^+ phage have an increased plaque size on *recBCD*$^+$ hosts (McMilin *et al.*, 1974; Lam *et al.*, 1974; Henderson & Weil, 1975). Approximately 100 large plaques were picked into phage buffer, and the eluted phage spotted onto bacterial lawns of 594 (*rec*$^+$), N2364 (*sbcC*) and JC9387 (*recBC sbcBC*). Phage growing efficiently on N2364 and JC9387, but not on 594 were selected, and purified by two rounds of growth on JC9387. Single plaques were picked for the preparation of phage stocks by plate lysates, again using JC9387 as a host. The recombinant phage was (eventually) designated DRL186.

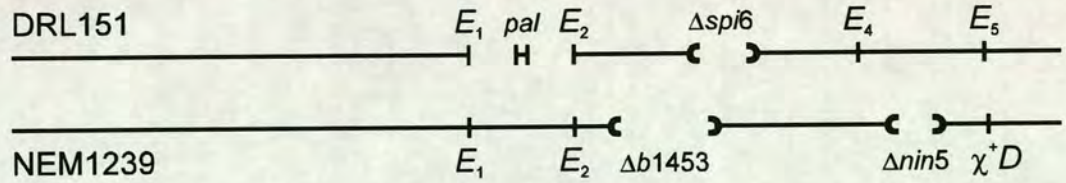
Location of the χ^+ Site in NEM1239 and DRL186

The χ^+ site in NEM1239 was reportedly χ^+D, and should therefore be situated at bp 45025– 45032 (Smith *et al.*, 1981) of the wild-type λ sequence (Sanger *et al.*, 1982). In order to localise the crossover point of the recombinant phage, λ DNA was prepared from the plate lysates and cleaved to completion with *Eco*RI. NEM1239 is an *srI4*° *srI5*° λ derivative and therefore lacks *Eco*RI sites 4 and 5 (bp 39168 and 44972 of the wild-type λ sequence respectively). The absence of *Eco*RI sites 4 and 5 and presence of a χ^+ site in a DRL151 derivative had been the desired outcome of the cross. However, the experiments calling for a palindrome-bearing phage without *Eco*RI sites 4 and 5 were not performed, and are consequently not described here.

The *Eco*RI restriction digest showed that the recombinant phage had retained *Eco*RI sites 4 and 5, and furthermore produced *Eco*RI restriction fragment sizes identical to the parental DRL151. This implies that the crossover had occurred between *Eco*RI site 5 (bp 44972) and χ^+D (bp 45025). The λ maps in Fig 5.1 demonstrate clearly that the likelihood of this event is very remote.

Figure 5.1

Colinear λ maps of DRL151 and NEM1239.



The theoretical probability that a crossover occurring in the region from the palindrome to χ^+D falls in the 47 bp interval between *Eco*RI site 5 and χ^+D is 5×10^{-3}.

Of the 100 phage that had been selected for the presence of a χ^+ site by plating on an *sbcC* host, three failed to grow on an *sbcC*$^+$ strain and were therefore assumed to be carrying the palindrome. As all three of these recombinants had retained *Eco*RI sites 4 and 5, it is virtually impossible that the crossovers had occurred between *Eco*RI site 5 and χ^+D.

The best explanation for these observations is that the χ^+ site in NEM1239 is not χ^+D, but lies to the left of *Eco*RI site 1; χ^+A is the only known site in this region. A crossover occurring in the left arm of DRL151 would account for the observed linkage of the palindrome and *Eco*RI sites 4 and 5. Unfortunately χ^+A is the only χ^+ site of λ that has not been located by DNA sequence analysis. An attempt to redress this disparity is presented below.

DNA Sequence Analysis of Potential χ^+A Sites in DRL186

Initially, the absence of a χ^+D site in both NEM1239 and DRL186 was determined by bidirectional DNA sequence analysis. Dideoxynucleotide chain-terminating reactions (Sanger *et al.*, 1977) were performed using λ NEM1239 and DRL186 DNA that had been prepared from a phage liquid lysis, and primers complementary to the λ sequence on the left and right of χ^+D at bp 45025– 45032. Neither phage differed from the wild-type sequence (5' GCGGGTGG 3') in this region, confirming that the χ^+ site in NEM1239 and DRL186 is not χ^+D.

Table 5.1

Location and sequence of potential χ^+A mutations in λ (see text for details).

Location in wild-type λ	Sequence of *I* strand in λ
bp 15166 – 15173	GC**C**GGTGG → GC**T**GGTGG
bp 15712 – 15719	GCTGGTG**A** → GCTGGTG**G**
bp 15864 – 15871	**A**CTGGTGG → **G**CTGGTGG
bp 16164 – 16171	GC**A**GGTGG → GC**T**GGTGG

Potential χ^+A sites were established by analysis of the wild-type λ sequence (Sanger *et al.*, 1982). χ^+A has previously been mapped genetically to a region of λ encompassing the *I* gene and the left ~14% of the *J* gene (G. Smith, personal communication); this corresponds to the interval bp 14773 ~ 15980. All χ^+ mutations that have been analysed at a DNA sequence level involve a single base pair change

(Smith, 1988b). If χ^+A is similarly the result of a single base pair mutation, then there are three potential χ^+A sites in the *I* – 14% *J* region. The locations of the sites are shown in Table 5.1, in addition to their sequence in the *l* strand of λ (the strand to which leftward-directed transcripts hybridise). The sequence of χ^+ is shown to the right of each potential site. A fourth site at 19.5% *J* is also shown.

The nucleotide sequence of DRL186 at each of these four potential χ^+A sites was determined by dideoxynucleotide chain-terminating reactions, but in each case was found to be no different to the wild-type λ sequence. If χ^+A lies in the *I* – 14% *J* interval, then the absence of a consensus χ^+ site in DRL186 at any of the locations listed in Table 5.1 may be due to one of two reasons:

1. the χ^+A mutation is the result of an insertion or deletion, and not the result of a single base pair change,
2. or the χ^+ site in the left arm of DRL186 is not a genuine χ^+A mutation.

The most effective way of differentiating between these two possibilities would be to determine the nucleotide sequence of a genuine χ^+A mutant at the sites listed in Table 5.2. This would make efficient use of the sequencing primers that had been used here.

Analysis of *in Vivo* Dam Methylation of DRL186

In Vivo Methylation by Growth of DRL186 in N2364 and JC7623 Hosts

DRL186 was used to infect N2364 (*sbcC dam*$^+$) and JC7623 (*recBC sbcBC dam*$^+$) at a MOI of 0.1, and grown by liquid lysis. After complete cell lysis, phage particles were purified and λ DNA extracted using the large scale method. JC7623 was used as a *recBC sbcBC* host instead of JC9387, as JC7623 and N2364 constitute an isogenic pair, which differ only in the presence of *recBC sbcB* mutations (see Table 2.1); the *sbcC* mutation in N2364 had been introduced by means of a transduction from JC7623 (Lloyd & Buckman, 1985).

Analysis of Methylation at GATC Sites in DRL186

The extent of Dam methylation at the palindrome centre was assayed by *Dpn*I and *Mbo*I cleavage as described in Chapter 4. 2 μg of DRL186 λ DNA isolated after growth in N2364 and JC7623 was initially digested to completion with *Eco*RI and radiolabelled using Klenow enzyme. [α^{35}S]dATP was used instead of [α^{32}P]dATP, primarily for health and safety reasons; β particles emitted by ^{35}S have a lower average energy ($\overline{E}_\beta = 0.0492$MeV) than those emitted by ^{32}P ($\overline{E}_\beta = 0.70$MeV). The DNA was

then extracted with phenol-chloroform and chloroform, precipitated with ethanol, resuspended in buffer and digested with *Bam*HI, *Dpn*I or *Mbo*I, or left uncut, as described in Chapter 4. Following a second round of phenol-chloroform and chloroform extraction and ethanol precipitation, the DNA was resuspended in gel-loading buffer and electrophoresed as described previously. The dried gel was autoradiographed to provide the image in Fig. 5.2, or exposed to a storage phosphor screen for the quantification of band intensities.

Quantification of Methylation at GATC Sites

Quantification was performed using a PhosphorImager™, which uses storage phosphor to detect radioactive disintegration. This system reportedly has a linear response to β particle emissions over a 10^5-fold range of intensities, whereas X-ray film has at best a semi-linear response over a 10^2-fold range. The bands labelled in Fig. 5.2 were quantified and their values are shown in Tables 5.2 and 5.3. The degree of Dam methylation at various GATC sites *in vivo* and the relative efficiency of cleavage by *Dpn*I and *Mbo*I at these sites were calculated as for Table 4.1, and are also shown in Tables 5.2 and 5.3. The latter set of values are plotted in Figs. 5.3 and 5.4.

Detection of Radiolabelled 416 bp *cos* LE – GATC Band

The autoradiograph in Fig. 5.2 shows a number of doublet bands that are not apparent in Fig. 4.4. The higher resolution of Fig. 5.5 is due to the low average energy of β particles emitted by ^{35}S isotopes, which penetrate only the film emulsion layer closest to the dried gel. On the other hand, β particles emitted by ^{32}P isotopes penetrate the film base and the second layer of emulsion of double-sided X-ray film, thereby resulting in a blurred image.

The doublet band labelled Palindrome in the rightmost lane is the result of hairpin formation by the full-length palindrome fragment. The self-annealed molecules migrate more rapidly during gel electrophoresis, and their formation is favoured at the edges of the gel due to lower temperature. The origin of the doublet band labelled Half Palindrome in the lanes corresponding to *Eco*RI – *Bam*HI digestion is not known, but is of no consequence to the methylation assay. However, the band previously labelled 409 bp Control consists of a doublet in both *Dpn*I and *Mbo*I digests, and the unwitting quantification of both bands may have led to the non-reciprocity in *Dpn*I and *Mbo*I cleavage efficiencies previously observed at this site (see Chapter 4). The upper band (labelled 416 bp *cos* LE) results from radiolabelling of the cohesive left end (*cos* LE) of λ and cleavage of the adjacent GATC site at bp 416 (Sanger *et al.*, 1982).

Figure 5.2

Autoradiograph of Restriction Digests at GATC Sites in λ DRL186

Autoradiograph showing restriction digests of λ DRL186 DNA determining the level of Dam methylation at various GATC sites *in vivo*. Two sets of digests are shown, representing methylation in different *E. coli* hosts. The λ phage had been grown in either N2364 (*sbcC*) or JC7623 (*recBC sbcBC*) and DNA purified as described in the text. The DNA was cleaved with *Eco*RI and radiolabelled with [α^{35}S]dATP as described in the text, before cleavage with *Bam*HI, *Dpn*I or *Mbo*I. The *E. coli* host strain and the appropriate enzymes are indicated above each lane. The full-length and half-length palindrome, and the control fragments (see Fig. 4.3) are shown by the labels on the right of the autoradiograph. For a detailed explanation see Fig. 4.4. The 416 bp DNA fragment resulting from radiolabelling of the *cos* left end of λ and cleavage at the adjacent GATC site is also shown.

From left to right, the lanes are:

1. λ DRL186 [N2364 (*sbcC*) host] *Eco*RI
2. λ DRL186 [N2364 (*sbcC*) host] *Eco*RI *Bam*HI
3. λ DRL186 [N2364 (*sbcC*) host] *Eco*RI *Dpn*I
4. λ DRL186 [N2364 (*sbcC*) host] *Eco*RI *Mbo*I
5. λ DRL186 [JC7623 (*recBC sbcBC*) host] *Eco*RI
6. λ DRL186 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Bam*HI
7. λ DRL186 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Dpn*I
8. λ DRL186 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Mbo*I

Digestion with only *Eco*RI generates the full length palindrome, while additional cleavage by *Bam*HI at the central GGATCC site generates the half-length palindrome. Digestion with *Dpn*I indicates the level of fully methylated DNA, and *Mbo*I the level of fully unmethylated DNA.

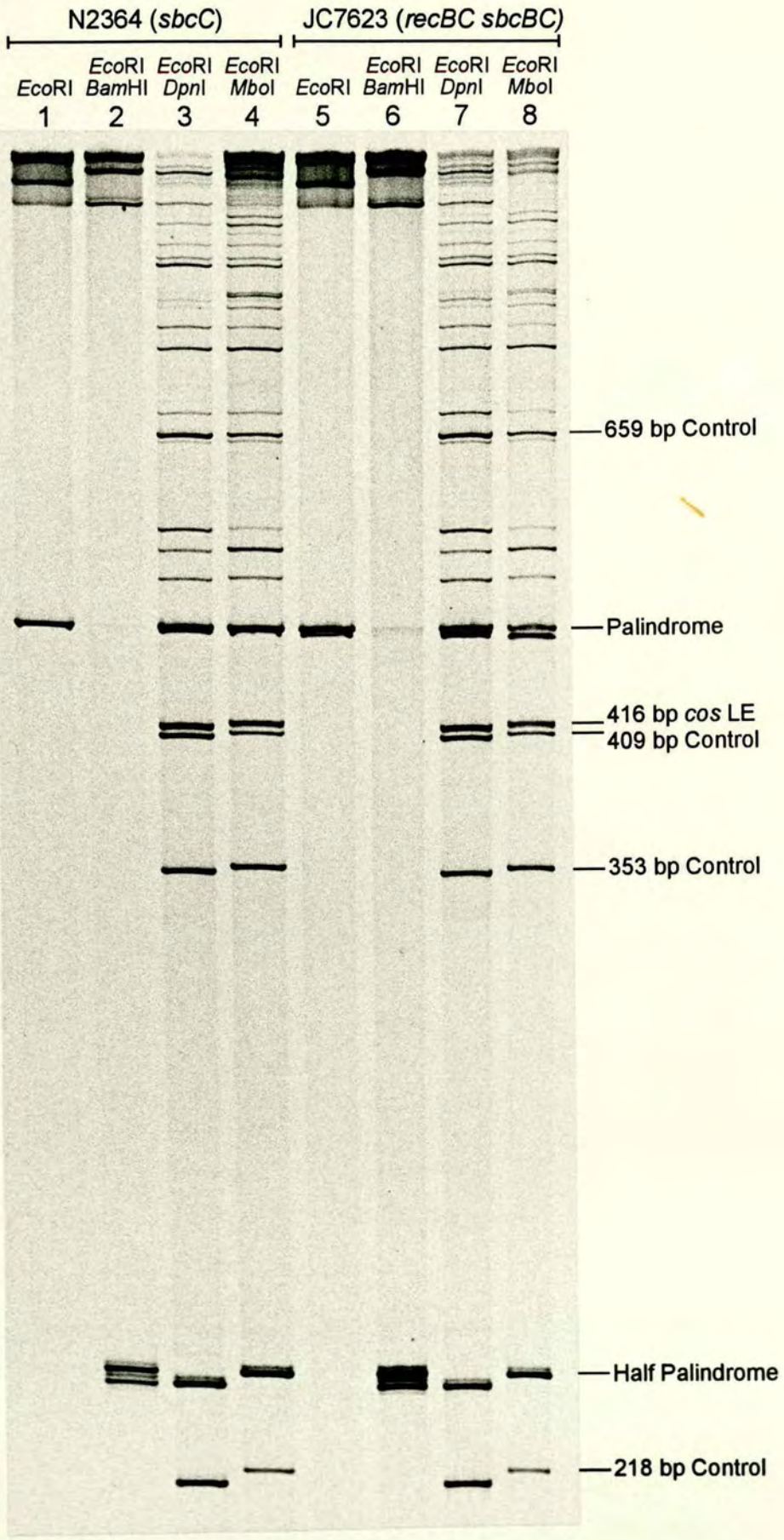

N2364 (sbcC)
JC7623 (recBC sbcBC)
EcoRI
EcoRI BamHI
EcoRI DpnI
EcoRI MboI
EcoRI
EcoRI BamHI
EcoRI DpnI
EcoRI MboI
1
2
3
4
5
6
7
8
659 bp Control
Palindrome
416 bp cos LE
409 bp Control
353 bp Control
Half Palindrome
218 bp Control

Table 5.2

Methylation in N2364 (*sbcC*) of GATC Sites in DRL186

The bands shown in the four leftmost lanes of Fig. 5.2, which correspond to methylation in N2364 (*sbcC*), were quantified using a PhosphorImager™. Their intensities are displayed in Table 5.2 in the rows designated "Data". The quantified values (Uncorrected Data) for the full-length and half-length palindrome were halved (Corrected Data). The % cleavage at the palindrome centre by *Dpn*I or *Mbo*I was calculated using equation (9) and the % methylation at the palindrome centre was determined using the level of methylated DNA as measured by both *Dpn*I and *Mbo*I cleavage; the values from the two enzymes were averaged using equation (10). The average % cleavage at control sites was determined by equation (11) using the sum of the palindrome bands as a standard, and the average % methylation of the control sites was similarly calculated using equation (12). The bottom portion of the table shows the relative efficiency of cleavage by *Dpn*I and *Mbo*I at the palindrome and control GATC sites. For a detailed explanation of these calculations see Table 4.1

Table 5.3

Methylation in JC7623 (*recBC sbcBC*) of GATC Sites in DRL186

The bands shown in the four rightmost lanes of Fig. 5.2, which correspond to methylation in JC7623 (*recBC sbcBC*), were quantified using a PhosphorImager™. Their intensities are displayed in Table 5.3 in the rows designated "Data". The values derived from these data were calculated as for Table 5.2 (see above).

Table 5.2

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$		*Mbo*I (GATC)
Uncorrected Data	Palindrome	6462		8186
	½ Palindrome	3916		4679
Corrected Data	Palindrome	3231		4093
	½ Palindrome	1958		2340
% Cleavage at Palindrome Centre		37.7		36.4
% Methylation of Palindrome Centre			50.7	
Control Data	659 bp Control	2806		1260
	409 bp Control	3173		1057
	353 bp Control	2966		2330
	218 bp Control	2781		885
	Average Control	2932		1383
Average % Cleavage at Control Sites		56.5		21.5
Average % Methylation of Control Sites			67.5	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	66.8		169.2
	669 bp Control	95.7		91.1
	409 bp Control	108.2		76.4
	353 bp Control	101.2		168.5
	218 bp Control	94.9		64.0
	Average Control	100		100

Table 5.3

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$		*Mbo*I (GATC)
Uncorrected Data	Palindrome	7262		3639
	½ Palindrome	3608		3326
Corrected Data	Palindrome	3631		1820
	½ Palindrome	1804		1663
% Cleavage at Palindrome Centre		33.2		47.8
% Methylation of Palindrome Centre			42.7	
Control Data	659 bp Control	3865		893
	409 bp Control	2965		788
	353 bp Control	2464		1867
	218 bp Control	2895		835
	Average Control	3047		1096
Average % Cleavage at Control Sites		56.1		31.5
Average % Methylation of Control Sites			62.3	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	59.2		151.8
	669 bp Control	126.8		81.5
	409 bp Control	97.3		71.9
	353 bp Control	80.9		170.4
	218 bp Control	95.0		76.2
	Average Control	100		100

Figure 5.3

Relative Cleavage by *Dpn*I at GATC Sites in DRL186

Graph showing the relative efficiency of cleavage by *Dpn*I at GATC sites in λ DRL186 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨). The values plotted are those shown on the left-hand side of the bottom portions of Tables 5.2 and 5.3. The intensities of the bands shown in Fig. 5.2 are presented relative to the average intensity of the control bands, although the value for the half-length palindrome band was halved to correct for radiolabelling at two *Eco*RI sites. Error bars are also shown, representing the range of relative cleavage by *Dpn*I at GATC sites as calculated from individual restriction digests. Figure 5.3 represents graphically the presence of Dam methylation at various GATC sites in DRL186.

Figure 5.4

Relative Cleavage by *Mbo*I at GATC Sites in DRL186

Graph showing the relative efficiency of cleavage by *Mbo*I at GATC sites in λ DRL186 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨). The values plotted are those shown on the right-hand side of the bottom portions of Tables 5.2 and 5.3. The data are presented as in Fig. 5.3. Error bars are also shown, representing the range of relative cleavage by *Mbo*I at GATC sites as calculated from individual restriction digests. Figure 5.4 represents graphically the absence of Dam methylation at various GATC sites in DRL186.

Figure 5.3

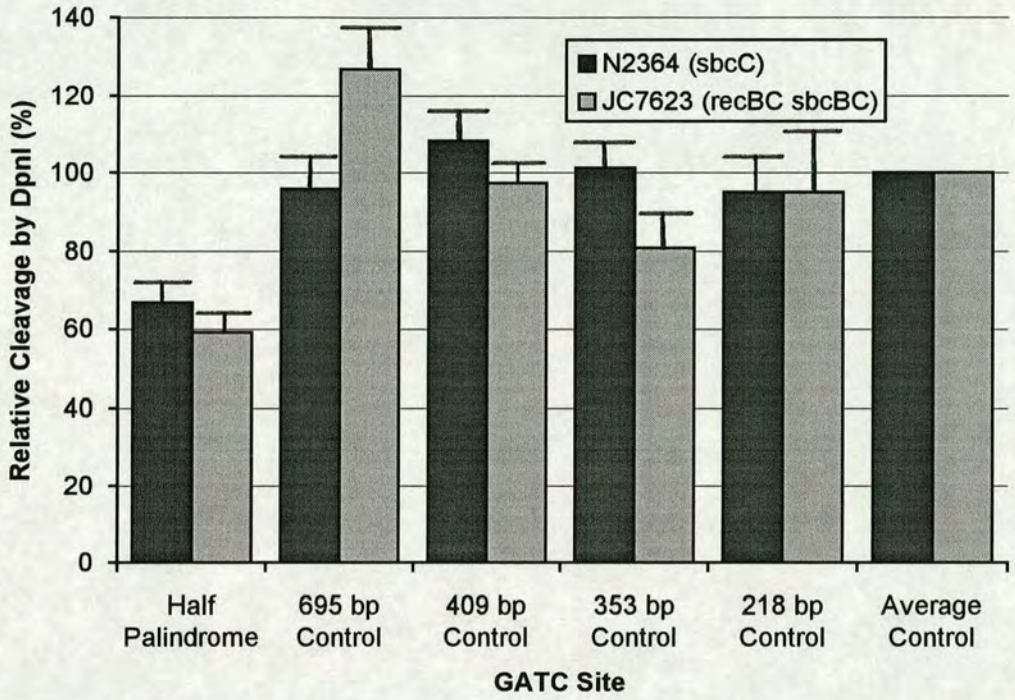
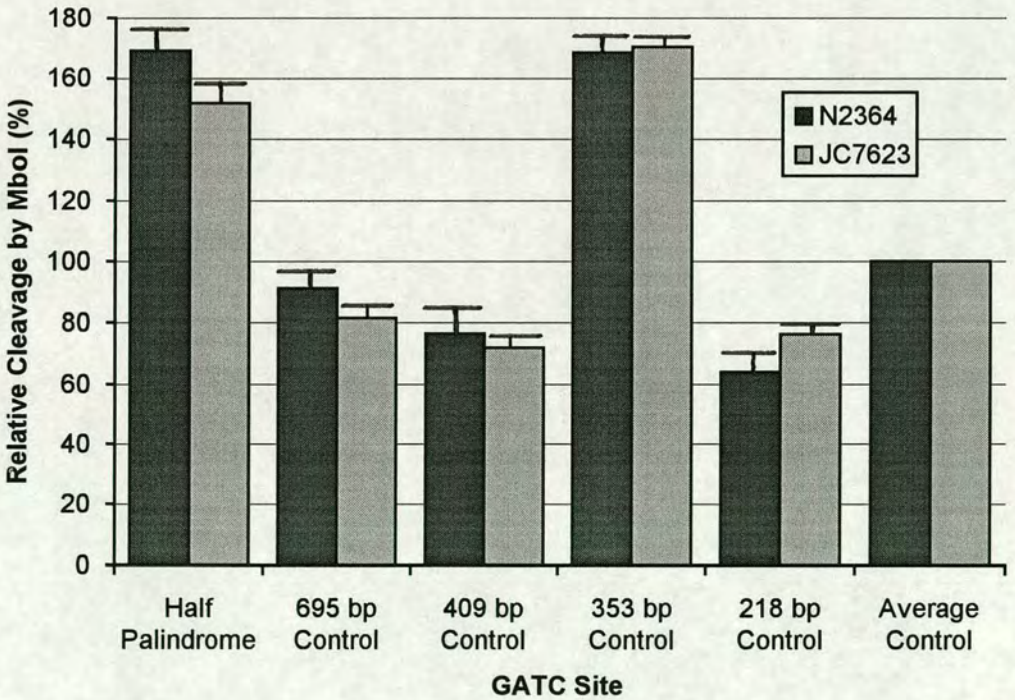


Figure 5.4



Conclusion

The results presented in Tables 5.2 and 5.3 show that the palindrome centre of DRL186 is undermethylated compared to the control GATC sites. However, it is not immediately clear from these data whether the palindrome centre exhibits greater undermethylation in N2364 or JC7623. The % methylation of the palindrome GATC site and average of the control GATC sites are therefore shown in Table 5.4 below.

Table 5.4

Comparison of Dam methylation at GATC sites in DRL151 and DRL186.

Bacteriophage λ Strain		DRL151	DRL186	
Escherichia coli Host		JC9387	N2364	JC7623
% Dam Methylation	Palindrome Centre	34.7	50.7	42.7
	Average Control	56.2	67.5	62.3
Palindrome / Control Ratio		0.62	0.75	0.69

Table 5.4 shows that the relative undermethylation of the palindrome centre, which is expressed as the Palindrome / Control Ratio, differs after growth in different *E. coli* hosts: DRL186 shows greater undermethylation of the palindrome centre in JC7623 than in N2364. However, it is unlikely that this increased undermethylation is a direct result of the *recBC* mutations in JC7623. For example, λ DRL151 exhibits greater undermethylation in JC9387 than DRL186 does in JC7623, although these phage are expected to behave identically in *recBC sbcBC* strains; the χ^+ site of DRL186 affords it no advantage in a *recBC sbcBC* host.

This difference between the palindrome undermethylation of DRL151 and DRL186 in *recBC sbcBC* cells is most probably due to differing *E. coli* host strains; DRL151 was grown in JC9387 whereas DRL186 was grown in JC7623. The use of a different host strain may affect the rate of λ replication and therefore the degree of Dam methylation. A relationship between the overall level of methylation and the relative undermethylation of the palindrome centre can be seen in Table 5.4. It is therefore likely that the variation in palindrome undermethylation is not the result of differing methylation inhibition at that GATC site, but is caused by an alteration of the global level of Dam methylation, which in turn changes the sensitivity of the assay (see Introduction to Chapter 3).

These conclusions are reinforced by Figs. 5.3 and 5.4. A side-by-side comparison of the efficiency of *Dpn*I and *Mbo*I cleavage at GATC sites in DRL186 shows little difference between phage that have been grown in N2364 and JC7623. This form of analysis takes into account overall levels of Dam methylation, and a visual inspection of Figs. 5.3 and 5.4 is sufficient to confirm that the presence of additional *recBC* mutations in JC7623 does not alter the undermethylation of the palindrome centre significantly. This may indicate that the switch from θ to rolling circle λ replication does not affect inhibition of methylation at the palindrome centre. However, χ sites have been shown to protect against RecBCD-mediated degradation of double-stranded DNA tails produced during rolling circle plasmid replication (Dabert *et al.*, 1992). By analogy, the principal role of χ in the survival of *red gam* λ such as DRL186 may be to protect rolling circle concatemers from degradation by RecBCD. The similarity in levels of palindrome centre undermethylation observed in N2364 and JC7623 may therefore by due to (χ-facilitated) rolling circle replication in both hosts.

Figs. 5.3 and 5.4 also illustrate two additional points of interest. Firstly, the use of [α^{35}S]dATP has enabled separate quantification of the two bands in Fig. 5.2 labelled 409 bp Control and 416 bp *cos* LE, and therefore the accurate measurement of Dam methylation at the control GATC site corresponding to the 409 bp band. The expected reciprocity in *Dpn*I and *Mbo*I cleavage efficiency is consequently observed at this site (compare Figs. 5.3 and 5.4, and Figs. 4.5 and 4.6). Secondly, the control GATC site identified by the 353 bp band is cleaved very efficiently by *Mbo*I, which may indicate undermethylation of the control target *in vivo*. The proximity of this GATC site to the λ origin of replication lends credence to this idea; persistent binding by λ O protein would prevent Dam methylation. However, the near-average cleavage of this GATC site by *Dpn*I indicates that the increased efficiency of *Mbo*I digestion is more probably a result of site-specificity by the latter enzyme than inhibition of Dam methylation *in vivo*. It is therefore imperative to determine the effect of sequence context on Dam methylase activity and *Dpn*I and *Mbo*I cleavage. This is described in Chapter 6.

In summary, this chapter shows that the presence of a χ^+ site in λ, and the consequent growth of this phage in a *recBCD*$^+$ (*sbcC*) host does not significantly alter the undermethylation of the palindrome centre *in vivo*.

CHAPTER 6

ANALYSIS OF *IN VITRO* DAM METHYLATION AT GATC SITES

Introduction

Targets sites for DNA restriction and modification enzymes differ in their efficiency of cleavage or modification. The influence of DNA sequence context on cognate interactions between enzymes and their consensus DNA targets can be profound. It is therefore important to consider whether the undermethylation observed at the palindrome centre of DRL151 and DRL186 is the result of methylation inhibition by a cruciform structure, or merely due to the sequence context of that GATC site.

In order to assess the influence of the sequence context of palindrome and control GATC sites on Dam methylation, the λ DNA is modified *in vitro* under conditions where cruciform structures are unlikely to form. As cruciform extrusion is thermodynamically unfavourable in relaxed DNA, this provision is easily met by ensuring that the DNA is in a linear, relaxed conformation. A comparison of *in vivo* and *in vitro* Dam methylation efficiency at GATC sites may also reveal differences that are due to supercoiling-dependent DNA secondary structures *in vivo*. Similar *in vivo* studies of unusual DNA secondary structure have also used *in vitro* modification of relaxed DNA as a control experiment for the effect of sequence context on methylation inhibition (e.g. Jaworski *et al.*, 1987).

Results

Generation of Unmethylated λ DNA

DRL151 and DRL186 were used to infect JL32 (*recBC sbcBC dam*) at a MOI of 0.1, and grown by liquid lysis. After complete cell lysis, phage particles were purified and λ DNA extracted using the large scale method. JL32 was first examined for the *dam* phenotype by virtue of its survival after UV-irradiation; *recBC sbcBC dam* strains show only 1.8% survival after UV-irradiation with 20 J m^{-2}, while *recBC sbcBC* cells show 32% survival (McGraw & Marinus, 1980). The UV sensitivity of JL32 was therefore compared to that of its *dam*$^+$ parental strain JC9387, and colonies were chosen on the basis of substantially increased mortality after 20 J m^{-2} irradiation. The λ DNA isolated after growth in JL32 was examined for the absence of Dam methylation at GATC sites by digestion with *Dpn*I or *Sau*3AI. Only *Dpn*I was unable to cleave the λ DNA, confirming that it was unmethylated.

Analysis of *in Vitro* Methylation by Dam Methylase

In Vitro Methylation using Dam Methylase

λ DNA extracted from phage particles is linear and therefore relaxed. The λ DNA was first digested to completion with *Eco*RI, then radiolabelled using Klenow enzyme and either [α^{32}P]dATP or [α^{35}S]dATP. After phenol-chloroform and chloroform extraction and ethanol precipitation, the DNA was resuspended in the appropriate buffer and methylated using 6 U Dam methylase per μg λ DNA and 80 μM S-adenosyl methionine (SAM) for various times ranging between 0 and 100 minutes at 37°C.

Loss of Palindrome DNA Fragments at High Temperature

The *in vitro* Dam methylation reaction was initially stopped by heat denaturation of the enzyme at 65°C for 10 minutes. When the DNA was subsequently analysed for methylation at GATC sites, the full-length and half-length palindrome fragments were not recovered efficiently. Furthermore, this selective loss of the palindrome appeared to be exacerbated by increasing Dam methylation.

The methylation of both adenine residues in a GATC site lowers the average stability value (T_{MN}) (Gotoh & Tagashira, 1981) of the tetranucleotide from 76.63°C to 66.63°C (Collins & Myers, 1987); hemimethylated GATC sequences have a T_{MN} of 71.63°C. Local helix destabilisation by adenine (Dam) methylation at a palindrome centre has also been found to increase the rate of cruciform extrusion *in vitro* (Murchie & Lilley, 1989).

The palindromic DNA studied in the *in vitro* methylation assay described here is relaxed, and therefore unlikely to undergo cruciform extrusion. Nevertheless, two hairpins may be formed by the melting of the palindromic sequence and the self-association of the single strands. This reaction would be significantly favoured by adenine methylation at the palindrome centre, which would reduce the temperature needed to melt that part of the sequence by 10°C. It is therefore possible that the termination of *in vitro* methylation by heating to 65°C resulted in hairpin formation by palindromic fragments that were methylated at the centre. This would not only result in the selective loss of that DNA fragment, but also in the underestimation of Dam methylation at that site by subsequent analysis.

The *in vitro* methylation reactions were therefore stopped by placing the samples on dry ice. This amended procedure almost completely eliminated the

selective loss of palindromic fragments. Experiments were also undertaken to determine whether the absence of NaCl in the Dam methylase buffer contributes to hairpin formation and palindrome loss. As increased ionic strength elevates the melting temperature of DNA, the *in vitro* methylation was performed in Dam methylase buffer that had been adjusted to 100 mM NaCl. The subsequent analysis using *Dpn*I and *Mbo*I cleavage is shown in Fig. 6.1. The quantified values are not presented here, but indicate that recovery of the palindrome fragment is unaffected by the NaCl concentration of the Dam methylase buffer. However, the efficiency of methylation was found to be significantly lower in 100 mM NaCl; Dam methylase is inhibited by ionic strengths of greater than 200 mM (Herman & Modrich, 1982).

Analysis of *in Vitro* Methylation at GATC Sites

The methylated DNA samples were extracted with phenol-chloroform and chloroform, precipitated with ethanol and resuspended in TE buffer. The extent of Dam methylation at the palindrome centre and control GATC sites was assayed by *Dpn*I and *Mbo*I cleavage as described in Chapter 4. Following a second round of phenol-chloroform and chloroform extraction and ethanol precipitation, the DNA was resuspended in gel-loading buffer and electrophoresed as described previously. The dried gel was autoradiographed to provide the image in Fig. 6.1, or exposed to a storage phosphor screen for the quantification of band intensities.

Quantification of Methylation at GATC Sites

The bands in lanes 11 and 12 of Fig. 6.1 were quantified by PhosphorImager™ and their values are shown in Table 6.1. These lanes show *Dpn*I and *Mbo*I digestion of ^{35}S-labelled DRL186 DNA which had been methylated *in vitro* for 60 minutes using standard (0 mM NaCl) Dam methylase buffer. This level of methylation corresponds reasonably to that observed in phage isolated from *dam*$^+$ hosts, and was therefore chosen for a side-by-side comparison with *in vivo* methylation data. The relative efficiency of cleavage by *Dpn*I and *Mbo*I of DRL186 DNA that had been methylated *in vitro* using purified Dam methylase, and *in vivo* by growth in N2364 and JC7623 is shown in Figs. 6.2 and 6.3. These data are discussed in the Conclusion to this chapter.

Figure 6.1

Autoradiograph of Restriction Digest at GATC Sites in λ DRL186 Following *in Vitro* Dam Methylation

Autoradiograph showing restriction digest of λ DRL186 DNA determining the level of *in vitro* Dam methylation at various GATC sites. The phage had been grown in JL32 (*recBC sbcBC dam*) and unmethylated DNA purified as described in the text. The DNA was cleaved with *Eco*RI and radiolabelled with [α^{35}S]dATP as described in the text, before *in vitro* methylation using purified Dam methylase and SAM. The reaction with Dam methylase was performed in standard Dam methylase buffer (0 mM NaCl) or in buffer that had been adjusted to 100 mM NaCl. The methylation was followed by cleavage with *Bam*HI, *Dpn*I or *Mbo*I as described previously. The time of incubation with Dam methylase, the concentration of NaCl in the methylation buffer, and the appropriate restriction enzymes are indicated above each lane. The full-length and half-length palindrome, and the control fragments (see Fig. 4.3) are shown by the labels on the right of the autoradiograph.

From left to right, the lanes are:

1. λ DRL186 [0 minutes Dam methylation] *Eco*RI *Bam*HI
2. λ DRL186 [0 minutes Dam methylation] *Eco*RI *Dpn*I
3. λ DRL186 [0 minutes Dam methylation] *Eco*RI *Mbo*I
4. λ DRL186 [10 minutes Dam methylation, 0 mM NaCl] *Eco*RI *Bam*HI
5. λ DRL186 [10 minutes Dam methylation, 0 mM NaCl] *Eco*RI *Dpn*I
6. λ DRL186 [10 minutes Dam methylation, 0 mM NaCl] *Eco*RI *Mbo*I
7. λ DRL186 [10 minutes Dam methylation, 100 mM NaCl] *Eco*RI *Bam*HI
8. λ DRL186 [10 minutes Dam methylation, 100 mM NaCl] *Eco*RI *Dpn*I
9. λ DRL186 [10 minutes Dam methylation, 100 mM NaCl] *Eco*RI *Mbo*I
10. λ DRL186 [60 minutes Dam methylation, 0 mM NaCl] *Eco*RI *Bam*HI
11. λ DRL186 [60 minutes Dam methylation, 0 mM NaCl] *Eco*RI *Dpn*I
12. λ DRL186 [60 minutes Dam methylation, 0 mM NaCl] *Eco*RI *Mbo*I
13. λ DRL186 [60 minutes Dam methylation, 100 mM NaCl] *Eco*RI *Bam*HI
14. λ DRL186 [60 minutes Dam methylation, 100 mM NaCl] *Eco*RI *Dpn*I
15. λ DRL186 [60 minutes Dam methylation, 100 mM NaCl] *Eco*RI *Mbo*I

Digestion with *Eco*RI and *Bam*HI generates the half-length palindrome. Digestion with *Dpn*I indicates the level of fully methylated DNA, and *Mbo*I the level of fully unmethylated DNA.

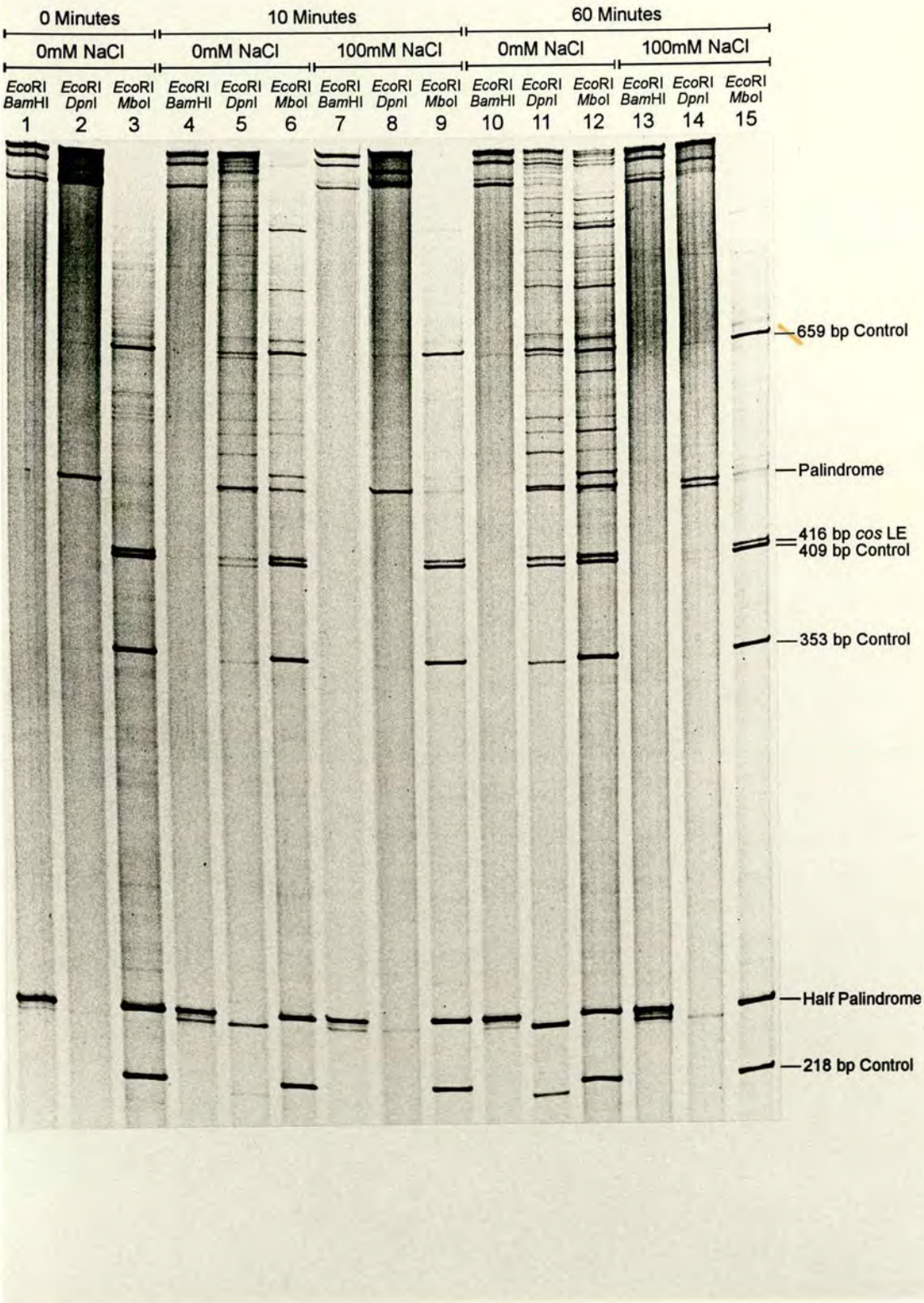

0 Minutes
10 Minutes
60 Minutes
0mM NaCl
0mM NaCl
100mM NaCl
0mM NaCl
100mM NaCl
EcoRI BamHI 1
EcoRI DpnI 2
EcoRI MboI 3
EcoRI BamHI 4
EcoRI DpnI 5
EcoRI MboI 6
EcoRI BamHI 7
EcoRI DpnI 8
EcoRI MboI 9
EcoRI BamHI 10
EcoRI DpnI 11
EcoRI MboI 12
EcoRI BamHI 13
EcoRI DpnI 14
EcoRI MboI 15
659 bp Control
Palindrome
416 bp cos LE
409 bp Control
353 bp Control
Half Palindrome
218 bp Control

Table 6.1

***In Vitro* Methylation of GATC Sites in DRL186**

The bands shown in lanes 11 and 12 of Fig. 6.1 were quantified by PhosphorImager™ and their values are displayed in Table 6.1 in the rows designated "Data". These lanes show *Dpn*I and *Mbo*I digestion of ^{35}S-labelled DRL186 DNA which had been methylated *in vitro* for 60 minutes using standard (0 mM NaCl) Dam methylase buffer. The quantified values (Uncorrected Data) for the full-length and half-length palindrome were halved (Corrected Data). The bottom portion of the table shows the relative efficiency of cleavage by *Dpn*I and *Mbo*I at the palindrome and control GATC sites.

Table 6.1

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$	*Mbo*I (GATC)
Uncorrected Data	Palindrome	5025	3411
	½ Palindrome	3048	3002
Corrected Data	Palindrome	2513	1706
	½ Palindrome	1524	1501
Control Data	659 bp Control	1809	3115
	409 bp Control	1215	3328
	353 bp Control	1259	6460
	218 bp Control	917	2538
	Average Control	1300	3860
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	117.2	38.9
	669 bp Control	139.1	80.7
	409 bp Control	93.5	86.2
	353 bp Control	96.8	167.4
	218 bp Control	70.6	65.8
	Average Control	100	100

Figure 6.2

Relative Cleavage by *Dpn*I at GATC Sites in DRL186

Graph showing the relative efficiency of cleavage by *Dpn*I at GATC sites in λ DRL186 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨), or *in vitro* using purified Dam methylase (■). The values plotted are those shown in Tables 5.2, 5.3 and 6.1 respectively. The intensities of the bands shown in Figs. 5.2 and 6.1 are presented relative to the average intensity of the control bands, although the value for the half-length palindrome band was halved to correct for radiolabelling at two *Eco*RI sites. Error bars are also shown, representing the range of relative cleavage by *Dpn*I at GATC sites as calculated from individual restriction digests. Figure 6.2 represents graphically the presence of Dam methylation at various GATC sites in DRL186.

Figure 6.3

Relative Cleavage by *Mbo*I at GATC Sites in DRL186

Graph showing the relative efficiency of cleavage by *Mbo*I at GATC sites in λ DRL186 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨), or *in vitro* using purified Dam methylase (■). The values plotted are those shown in Tables 5.2, 5.3 and 6.1 respectively. The data are presented as in Fig. 6.2. Error bars are also shown, representing the range of relative cleavage by *Mbo*I at GATC sites as calculated from individual restriction digests. Figure 6.3 represents graphically the absence of Dam methylation at various GATC sites in DRL186.

Figure 6.2

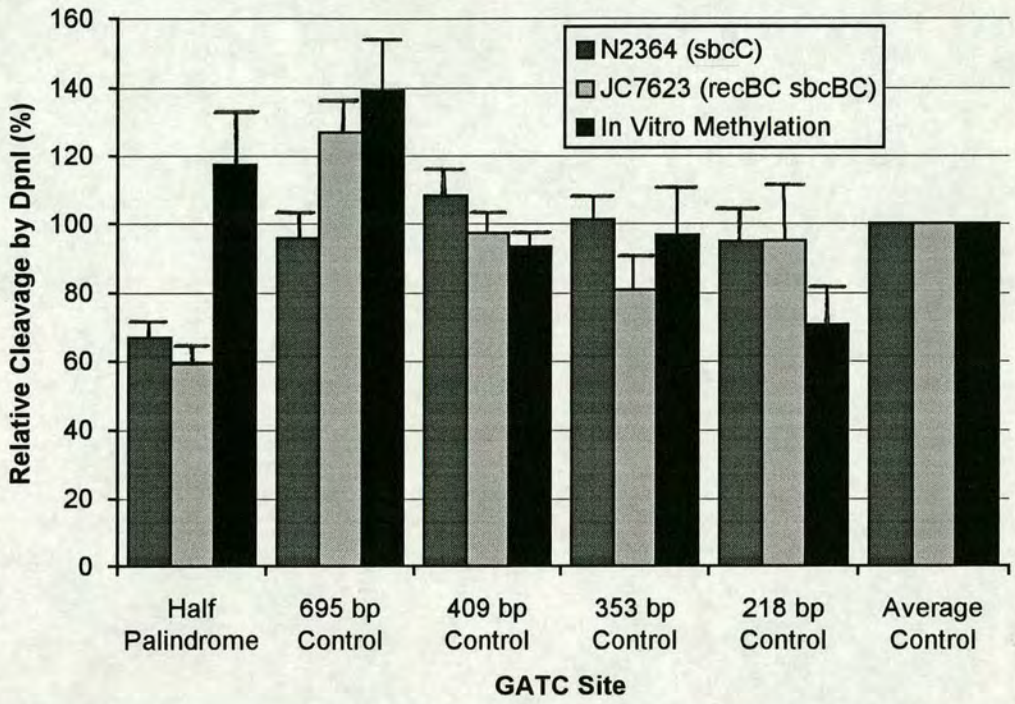
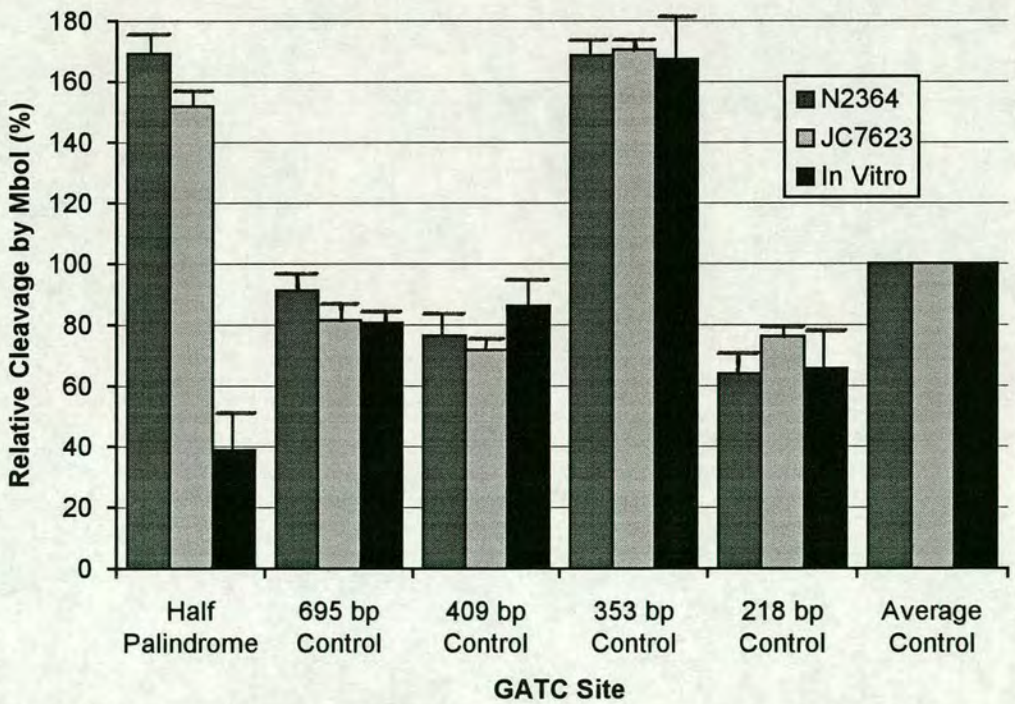


Figure 6.3



Conclusion

The results presented in this chapter are best illustrated by Figs. 6.2 and 6.3. It can be seen from these graphs that the GATC site at the palindrome centre is methylated with a high efficiency by Dam methylase *in vitro*. This is in stark contrast to the relatively low efficiency of methylation of this GATC target *in vivo*. Both *Dpn*I and *Mbo*I digests show a large difference in cleavage efficiency between a palindrome GATC site that has been methylated *in vivo*, and one methylated *in vitro*.

Conversely, the control GATC sites show a remarkable consistency of methylation efficiency between the *in vivo* and *in vitro* data. This consistency is reflected in both the *Dpn*I and *Mbo*I restriction digests: the differences in cleavage efficiency between control sites methylated *in vivo* and those methylated *in vitro* are relatively small. Furthermore, the *in vitro* methylation data strengthens the hypothesis proposed in Chapter 5 that the control GATC site identified by the 353 bp band is cleaved with unusual efficiency by *Mbo*I. If the substantial level of *Mbo*I cleavage had been due to methylation inhibition at this target site by λ O protein *in vivo*, then a similar degree of *Mbo*I digestion would not be expected after *in vitro* methylation.

The *in vivo* and *in vitro* results presented side-by-side in Figs. 6.2 and 6.3 argue for specific undermethylation of the palindrome centre *in vivo*. The difference between the efficiencies of *in vivo* and *in vitro* Dam methylation at the palindrome centre is substantially greater than the differences observed at the control GATC sites. This specific undermethylation of the palindrome centre *in vivo* is statistically significant at a confidence level of 98% in the case of the *Dpn*I digest of DRL186 grown in N2364, and at a confidence level of 99% for the other three restriction digests of *in vivo*-methylated λ DNA. (These confidence levels were determined by a one-tailed Student's *t* test using the *in vivo* / *in vivo* methylation differences of the control GATC sites as a *t* distribution with three degrees of freedom.)

In summary, the control experiment described in this chapter shows clearly that the undermethylation of the palindrome centre is not due to an effect of sequence context on methylation inhibition. Furthermore, a side-by-side comparison of *in vivo* and *in vitro* methylation data reinforces the conclusion that the palindrome GATC site is undermethylated *in vivo*. However, it does not exclude the possibility that the serendipitous formation of a protein binding site, and not cruciform extrusion *in vivo*, is responsible for the inhibition of methylation at the palindrome centre. This is discussed in greater detail in Chapter 8.

CHAPTER 7

CLONING OF $(GC)_n$ REPEAT IN λ AND pBR322 – DELETION OF Z-DNA?

Introduction

Like cruciform DNA, left-handed Z-DNA is an unusual secondary DNA structure whose formation is favoured by negative supercoiling. It is formed by alternating purine-pyrimidine sequences such as $(GC)_n$ or $(AC)_n$ (but not $(AT)_n$, which forms cruciform structures). Formation of Z-DNA by such repeats has been detected *in vitro* using two-dimensional gel electrophoresis (Haniford & Pulleyblank, 1983; Wang *et al.*, 1983), anti-Z-DNA antibody binding (Rich *et al.*, 1984), chemical probing (Herr, 1985; Johnston & Rich, 1985), S1 nuclease cleavage (Singleton *et al.*, 1982), and DNA methylase and restriction endonuclease inhibition (Vardimon & Rich, 1984; Zacharias *et al.*, 1984). The use of DNA methylases as structural probes has been extended to the detection of Z-DNA formation *in vivo* (Jaworski *et al.*, 1987; Jaworski *et al.*, 1988); these *in vivo* studies have been discussed in the Introduction to this thesis and in Chapter 3.

The formation of Z-DNA *in vivo* by a $(GC)_n$ repeat was chosen as a positive control experiment for the assay developed here due to four principal reasons:

1. Z-DNA has been detected *in vivo* using a methylase probe (Jaworski *et al.*, 1987),
2. like cruciform extrusion, Z-DNA formation is induced *in vitro* by supercoiling under physiological ionic conditions (Singleton *et al.*, 1982),
3. both cruciform extrusion by $(A\text{-}T)_n$ sequences (McClellan *et al.*, 1990; Zheng *et al.*, 1991; Dayn *et al.*, 1992) and Z-DNA formation (Rahmouni & Wells, 1989; Zheng *et al.*, 1991; Rahmouni & Wells, 1992) are driven by negative supercoiling *in vivo*,
4. unlike cruciform extrusion by palindromes of a normal base composition (Courey & Wang, 1983), Z-DNA formation is relatively rapid because of its significantly smaller kinetic barrier (Rich, 1984; Frank-Kamenetskii & Vologodskii, 1984).

Results

The cloning of a $(GC)_n$ repeat in bacteriophage λ was attempted. Due to the difficulties encountered in this undertaking, the $(GC)_n$ sequence was cloned in plasmid pBR322. This construct was then used in an effort to detect Z-DNA formation *in vivo* using Dam methylase.

Attempt to Clone $(GC)_n$ Repeat in Bacteriophage λ

In order to construct a λ derivative carrying a $(GC)_n$ sequence, an attempt was undertaken to insert such a repeat into DRL112 (*spi*6 *cI*857). DRL112 is isogenic to DRL133 (which is described in Chapter 4), save that DRL112 is not Δ*B* (it has not lost the *Eco*RI B fragment) and does not contain the 462 bp palindrome.

$(GC)_n$ Oligonucleotide with Central GATC Site

The oligonucleotide shown below was chosen by three criteria:

1. a similar repeat has been shown to form Z-DNA *in vivo* (Jaworski *et al.*, 1987),
2. the oligonucleotide contains a central target site for Dam methylase, which also forms part of a site for *Bam*HI restriction endonuclease,
3. the oligonucleotide has 3' terminal *Sac*I linker sequences which will inactivate the *Sac*I target site upon insertion (see Chapter 4).

```
                    Dam
5'      (GC)13GGATCC(GC)13AGCT 3'
3' TCGA(CG)13CCTAGG(CG)13      5'
   SacI         BamHI        SacI
```

Cloning of $(GC)_n$ Oligonucleotide in DRL112

The cloning strategy adopted here was to replace the non-essential 1.1 kb *Sac*I-*Sac*I fragment (bp 24776 – 25881 of wild-type λ) in the left arm of DRL112 with the $(GC)_n$ oligonucleotide shown above. This location would be comparable to that occupied by the 462 bp palindrome in DRL133. DRL112 was therefore cleaved to completion with *Sac*I. The $(GC)_n$ oligonucleotide had been heated to 100°C and allowed to cool to room temperature slowly to favour the formation of double-stranded DNA (see Chapter 4). It was ligated with the DRL112 DNA at a 5:1 (oligonucleotide:λ) molar ratio of cohesive *Sac*I termini. The DNA ligase was inactivated at 70°C and the reaction allowed to cool slowly as before. The DNA was cleaved to completion for a second time with *Sac*I in order to select against phage without an oligonucleotide insertion. The DNA was precipitated with ethanol, resuspended in 5μl of TE buffer, and *in vitro* packaged to produce mature λ particles.

The *in vitro*-packaged phage were amplified by one round of plating on a JC9387 (*recBC sbcBC*) host. A *recBC* mutant is necessary to permit rolling circle replication by a *spi*⁻ (*red gam*) χ° λ phage such as DRL112, and JC9387 was chosen as a host to maintain the analogy with the experiments described in Chapters 4 and 5.

24 plaques were picked into phage buffer, purified by two further rounds of growth on JC9387, and used for the preparation of phage stocks from plate lysates, again using JC9387 as a host. λ DNA was prepared from the plate lysates and cut with *Sac*I to identify phage which had escaped earlier cleavage; only one phage recut with *Sac*I.

The remaining 23 phage were analysed for the presence of the novel *Bam*HI site in the $(GC)_n$ insertion. The DNA was digested to completion with *Bam*HI and *Xba*I; the *Xba*I digest was performed to generate DNA fragments that are easily discernible by agarose gel electrophoresis. The novel *Bam*HI site could be found in none of the 23 phage. Nevertheless, the digest showed clearly that the *Sac*I-*Sac*I fragment of DRL112 had been replaced in all 23 phage with a small (20 ~ 40 bp) insertion. Furthermore, the earlier restriction digest indicated that the *Sac*I target sites had been inactivated. The simplest explanation for these observations is that the $(GC)_n$ oligonucleotide had replaced the *Sac*I-*Sac*I fragment as planned, but had subsequently suffered a deletion of 20 ~ 40 bp at its centre. This would account for the variable size of the insertion and the absence of a novel *Bam*HI site in the 23 recombinant phage.

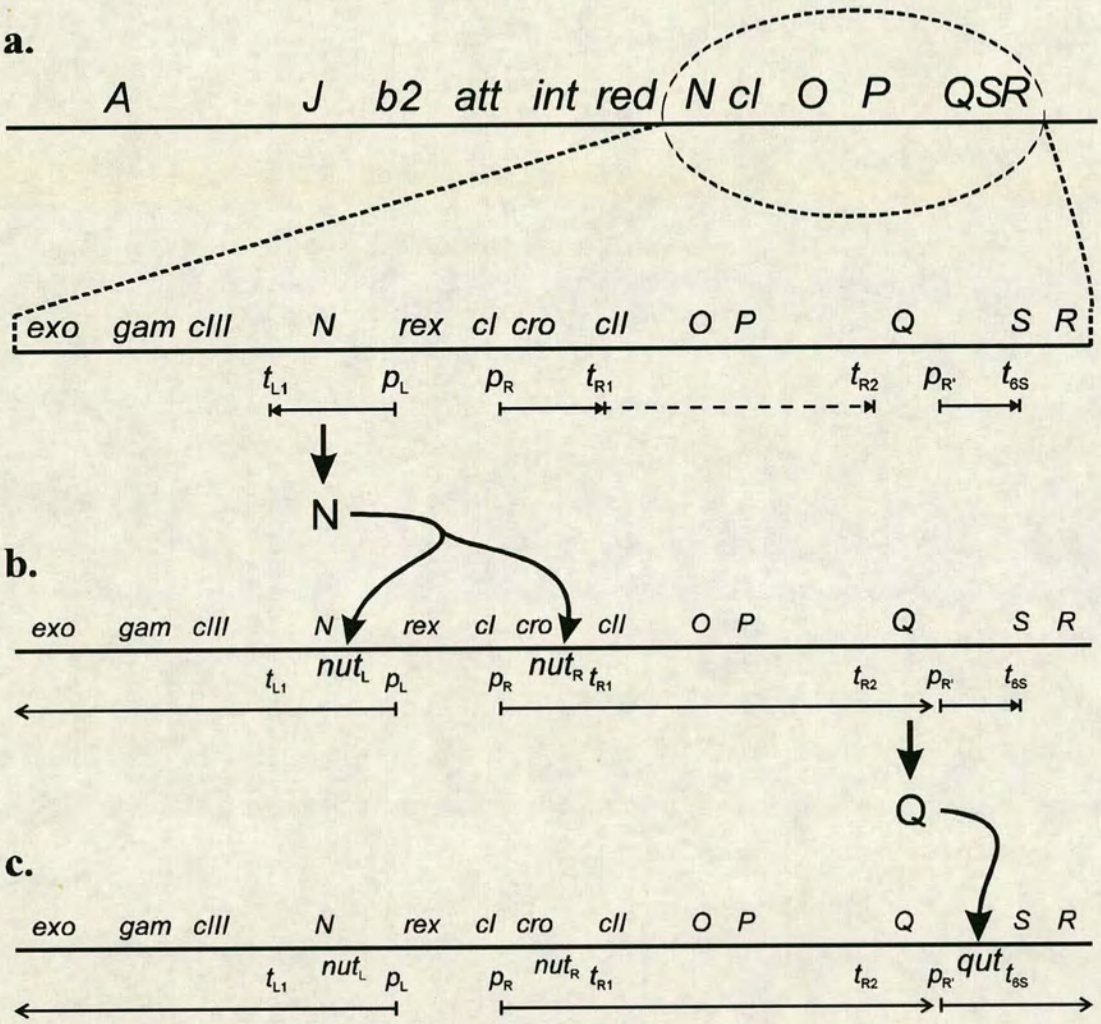
Cloning of $(GC)_n$ Oligonucleotide in NEM423 (N^-)

A possible reason for the deletion of the $(GC)_n$ oligonucleotide insertion is increased formation of Z-DNA as a result of supercoiling of the DNA template during transcription. Jaworski *et al.* (1989) found that while sequences capable of adopting a Z-DNA conformation were stable when cloned in the *Eco*RI site (an untranslated region) of pBR322, they suffered deletions when cloned in the *Bam*HI site, which is located in the tetracycline resistance structural gene (*tet*). However, when the promoter for the *tet* gene was inactivated, the inserts at the *Bam*HI site were not deleted. This was concomitant with a reduction in the negative superhelix density of the plasmid and the frequency of Z-DNA formation by these inserts.

These deletions are by implication the result of increased Z-DNA formation, which is due to elevated levels of negative supercoiling in transcriptionally active regions. Z-DNA formation by $(GC)_n$ tracts has been widely used as a probe of localised supercoiling (reviewed by Rahmouni, 1992) and has provided direct evidence for the effect of transcription on local DNA supercoiling *in vivo* (Rahmouni & Wells, 1992) as predicted by the twin supercoiling domain model of Liu & Wang (1987). The waves of supercoiling generated by the movement of the transcriptional apparatus favour the formation of Z-DNA, and therefore its deletion. The exact mechanism of Z-DNA deletion is not known, but may involve a subversion of cellular recombination events; this is discussed later.

Figure 7.1

Map of λ (not to scale) illustrating transcription patterns in the presence of N and Q (see text for details) (adapted from Friedman & Gottesman, 1983).



The deletion of the $(GC)_n$ insert during cloning in DRL112 may therefore have been caused by transcription through this region of λ, most probably originating from the p_L promoter. Unfortunately, transcriptional activity from p_L is relatively complex (see review by Friedman & Gottesman, 1983) and is therefore shown schematically in Fig. 7.1. Leftward transcription from p_L initially terminates at the t_{L1} terminator, producing only a transcript of the *N* gene (Fig. 7.1.a). However, *N* encodes an antitermination factor which interacts with RNA polymerase at *nut* sites enabling subsequent transcription from p_L to proceed through t_{L1}, and rightward transcription from p_R to proceed through t_{R1} and t_{R2} (Fig. 7.1.b). The p_R transcript encodes a second antitermination factor, the product of the *Q* gene, which acts on RNA polymerase at *qut* sites enabling rightward transcription originating at $p_{R'}$ to proceed through the t_{6S} terminator (Fig. 7.1.c).

Flanking the $(GC)_n$ insertion with transcriptional terminators would normally be expected to prevent deletion. However, the *N* gene product (N) is a powerful suppressor of all characterised rho-independent and rho-dependent terminators (Gottesman *et al.*, 1980; N. Franklin, personal communication). For example, N-mediated antitermination enables RNA polymerase to proceed through a tandem array of four powerful terminators with moderate efficiency (Franklin, 1989). Special terminators that are unresponsive to N-mediated antitermination would be required. Although such sites have been located in λ (Gottesman *et al.*, 1980; Burt & Brammar, 1982a), they have not been mapped to a nucleotide sequence level. The transcriptional isolation of the $(GC)_n$ repeat would therefore require the insertion of a fragment of λ spanning several kb (W. Szybalski, personal communication).

However, an alternative approach is possible: transcription from p_L may be prevented by a mutation in *N*, as the products of leftward transcripts are non-essential if provisions are made elsewhere in the λ and host genome. For instance, it is possible for λ to grow without a functional N protein if the terminators of essential rightward transcription are removed. This is accomplished by the *nin*5 deletion, which removes the t_{R2} terminator; t_{R1} is only 50% efficient (Rosenberg *et al.*, 1978) and therefore poses no serious problems in the absence of N-mediated antitermination.

λ NEM423 (*N cI*857 *nin*5) was therefore chosen as a vector for the $(GC)_n$ insert. The cloning was attempted as described above for DRL112: the NEM423 DNA was cut with *Sac*I, the oligonucleotide was ligated with the λ DNA at a 5:1 ratio of cohesive *Sac*I termini, digested for a second time with *Sac*I and *in vitro* packaged. The packaged phage were amplified by one round of growth on JC9387

(*recBC sbcBC sup*°). (N^- χ° phage such as NEM423 require host mutations in *recB* and *recC* for efficient plating as the genes transcribed from p_L include *red* and *gam*.) 30 plaques were picked into phage buffer, purified by two rounds of growth on JC9387, and used for the preparation of plate lysates.

λ DNA was prepared from the plate lysates and cut with *Sac*I. Seven phage which failed to cleave with *Sac*I were analysed for the presence of the $(GC)_n$ insertion by digestion with *Bam*HI. All seven phage had lost the 1.1 kb *Sac*I-*Sac*I fragment during cloning, but not one showed cleavage with *Bam*HI at this location. Instead, two distinct classes of deletion events could be deduced from the restriction digest: the first class involved a small (~50 bp) deletion of the oligonucleotide insertion site, whereas the second class consisted of larger deletion of ~2.5 kb including the oligonucleotide insertion site. The prevention of transcription from p_L by means of the *N* mutation had therefore failed to thwart the deletion of the $(GC)_n$ insertion.

Use of a *recA* Host Strain

As mentioned earlier, a possible mechanism for Z-DNA deletion may involve aberrant processing by the cellular recombination apparatus. Evidence supporting this hypothesis has mainly come from studies of Z-DNA binding by proteins involved in molecular synapsis during homologous recombination (reviewed by Holliday, 1989). The RecA protein of *E. coli* is stably bound by Z-DNA with an affinity 2 – 7 times greater than B-DNA (Blaho & Wells, 1987). The homologous pairing of plasmid DNA molecules by Rec1 protein of *Ustilago maydis*, which possesses virtually identical functions and reaction properties to RecA, is promoted by sequences of Z-DNA (Kmiec & Holloman, 1986).

It is therefore possible that deletion of the $(GC)_n$ insert in DRL112 and NEM423 was promoted by the host RecA protein. A *recA* host strain was therefore sought that would allow propagation of an *N* (*red gam*) χ° λ phage such as NEM423. Mutations in the host *recB* and *recC* genes normally allow efficient plating, but *recA recBC* mutants grow very poorly and are not good hosts for λ vectors. NM701 (*recA* p*gam*) was therefore used. The p*gam* plasmid (Crouse, 1985) expresses the λ *gam* gene under control of the λ $p_{R'}$ promoter, and serves to inactivate the host RecBCD protein during λ replication. The $p_{R'}$ promoter in p*gam* consists of a $p_{R'}$-*qut*-t_{6S} construct, and thereby prevents constitutive expression of *gam* in the absence of λ-encoded Q-mediated antitermination (see Fig. 7.1). The viability of a *red gam* χ° phage in NM701 was verified using λ MMS659 (*b*1453 *red gam* χ°).

However, NEM423 and its derivatives failed to grow in NM701. The most probable reason for this inviability is that insufficient Q protein is produced during infection by a *nin*5 phage. The *Q* gene is transcribed from the p_R promoter, and this transcription is normally antiterminated by N. An N^- mutant phage can grow if the t_{R2} terminator is deleted by *nin*5, as the t_{R1} terminator is only 50% efficient (Rosenberg *et al.*, 1978) (see Fig. 7.1). Although this results in a 50% reduction in levels of Q protein, it does not appear to cause problems for rightward transcription from λ $p_{R'}$. However, the Q protein is only significantly active when supplied in *cis* to its site of action (Burt & Brammar, 1982b). The 50% reduction in Q protein synthesis by an *N nin*5 phage such as NEM423 is therefore exacerbated by the *cis*-specificity of Q-mediated antitermination. Consequently, expression of the *gam* gene carried on the p*gam* plasmid is greatly reduced by virtue of decreased antitermination, and plating of NEM423 of NM701 is thus inefficient.

Perspectives – Use of *nut*L Phage

It is possible to prevent leftward transcription without compromising N-mediated antitermination of rightward transcription. This can be accomplished by a nut_L mutation which alters one of the sites of N-RNA polymerase interaction, thereby preventing the antitermination of p_L-initiated transcription (Salstrom & Szybalski, 1978). N-mediated antitermination at nut_R is unaffected, and rightward transcription therefore proceeds normally (see Fig. 7.1). A nut_L phage would be compatible with the p*gam* plasmid described above.

However, it is not certain that eliminating leftward transcription alone would prevent deletion of the $(GC)_n$ insert. During its lytic lifecycle, λ exists as a circular molecule and rightward transcription from $p_{R'}$ therefore continues into the left arm. It would be necessary to prevent this $p_{R'}$-initiated transcription from inducing deletion of the $(GC)_n$ sequence. Transcriptional termination sites that are unresponsive to Q-mediated antitermination have been identified in the λ *b*2 region (Burt & Brammar, 1982a). The *Sac*I sites of λ also lie in this region, and it therefore likely that $p_{R'}$-initiated transcription extends through the $(GC)_n$ insertion site and facilitates the formation of Z-DNA. It would therefore be more prudent to insert the $(GC)_n$ repeat to the right of these terminators. The third *Eco*RI site of λ at bp 31747 would constitute such a target site; it also lies to the left of t_{L1} and should therefore be transcriptionally silent in a nut_L phage.

Cloning of $(CG)_n$ Repeat in pBR322

Owing to the difficulties encountered in attempts to clone a $(GC)_n$ sequence in bacteriophage λ, pBR322 (Bolivar *et al.*, 1977b) was chosen an alternative vector. The stability of $(CG)_n$ inserts at the *Eco*RI site of this plasmid are well documented (Jaworski *et al.*, 1989), and the ability to perform such a construction successfully would therefore constitute a control experiment for the negative results obtained with λ vectors.

$(CG)_n$ Oligonucleotide

The oligonucleotide shown below was used for cloning in the *Eco*RI site of pBR322. It consists of two $(CG)_n$ repeats separated by a target site for Dam methylase, which also forms a *Bam*HI site. The oligonucleotide has 5' terminal *Eco*RI linker sequences, and is identical to the sequence in plasmid pRW478 which has been shown to form Z-DNA *in vivo* (Jaworski *et al.*, 1987).

```
                    Dam
5'  AATT(CG)13GATC(CG)13        3'
3'       (GC)13CTAG(GC)13TTAA  5'
   EcoRI       BamHI       EcoRI
```

Cloning of $(CG)_n$ Oligonucleotide in pBR322

The $(CG)_n$ oligonucleotide was phosphorylated using bacteriophage T4 polynucleotide kinase and ATP, extracted with phenol-chloroform and chloroform and precipitated with ethanol. It was resuspended in buffer, heated to 100°C and allowed to cool to room temperature slowly to favour the formation of double-stranded DNA (see Chapter 4). pBR322 was cleaved to completion with *Eco*RI, extracted with phenol-chloroform and chloroform and precipitated with ethanol. It was resuspended in the appropriate buffer and dephosphorylated using calf intestinal alkaline phosphatase (CIP) to prevent religation of the vector. The CIP was inactivated by heating to 75°C, and the DNA was extracted with phenol-chloroform and chloroform and precipitated with ethanol. The pBR322 DNA was resuspended in buffer and ligated with the $(CG)_n$ oligonucleotide at a 5:1 (oligonucleotide:plasmid) molar ratio of cohesive *Eco*RI termini. The DNA ligase was inactivated at 70°C and the reaction allowed to cool slowly. The ligation was used to transform DL655 (*recA sbcC*) to ampicillin and tetracycline resistance.

Single colonies resistant to ampicillin and tetracycline were purified and cultured in L broth supplemented with both antibiotics. Plasmid DNA was prepared

using the QIAGEN method and analysed for the presence of the $(CG)_n$ insert by a restriction digest with *Ssp*I and *Hin*dIII. The unique target sites for these two enzymes lie either side of the *Eco*RI site in pBR322, giving a fragment size of 222 bp if no insertion has occurred. After complete cleavage with *Ssp*I and *Hin*dIII, the DNA was radiolabelled with [α^{32}P]dCTP and Klenow enzyme, extracted with phenol-chloroform and chloroform, precipitated with ethanol, and resuspended in formamide-EDTA gel-loading buffer. Samples were electrophoresed on a 6% denaturing polyacrylamide gel, which was subsequently dried and autoradiographed.

Two plasmids containing the $(CG)_n$ insert were isolated by the restriction digest. Both plasmids yielded a 346 bp *Ssp*I-*Hin*dIII fragment, indicating that a multiple insertion of two 60 bp $(CG)_n$ sequences had occurred. The recombinant plasmid was designated pZTA1.

Stability of $(CG)_n$ Repeat in *sbcC* and *recA* Hosts

The stability of the $(CG)_n$ inserts in pZTA1 was investigated in a number of different host strains. As described earlier, interactions with RecA may be responsible for deletion of Z-DNA. The effect of the *sbcC* mutation was also analysed to determine whether the SbcCD protein can recognise more than one form of unusual DNA secondary structure.

pZTA1 was used to transform the strains shown in Table 7.1 to ampicillin resistance. Single colonies resistant to ampicillin were purified and cultured in L broth supplemented with antibiotic. Plasmid DNA was prepared using the small scale method and analysed for the presence of the $(CG)_n$ insert by a restriction digest with *Ssp*I and *Hin*dIII as described above.

Table 7.1

Relevant genotype of hosts strains used to determine $(CG)_n$ stability.

Host Strain	N2671	N2673	N2691	N2693
recA	+	+	–	–
sbcC	+	–	+	–

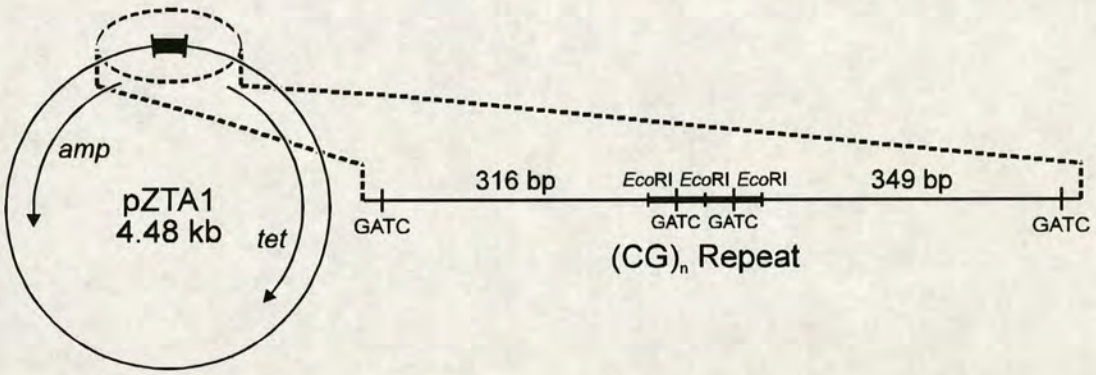
The restriction digest of pZTA1 DNA isolated from all four hosts showed clearly the presence of the 346 bp *Ssp*I-*Hin*dIII fragment. The tandem $(CG)_n$ repeat is therefore stable in all of these host strains, indicating that the effect of *recA* and *sbcC* mutations on Z-DNA deletion is negligible in this case.

Dam Methylation of $(CG)_n$ Repeat *in Vivo*

The extent of Dam methylation at GATC sites in the $(CG)_n$ inserts of pZTA1, and at two control GATC sites was assayed using *Dpn*I and *Mbo*I cleavage. The $(CG)_n$ repeats and the relevant control GATC sites (at bp 349 and 4047 of pBR322) are shown in the map of pZTA1 in Fig. 7.2 below.

Figure 7.2

Map of pZTA1 showing the $(CG)_n$ repeats (flanked by *Eco*RI sites) and the GATC sites assayed for Dam methylation.



pZTA1 DNA was prepared after growth in N2671, N2673, N2691 and N2693 (and DL655) as described above. As with bacteriophage λ, the plasmid DNA was first digested to completion with *Eco*RI and radiolabelled using [α^{35}S]dATP and Klenow enzyme. It was then extracted with phenol-chloroform and chloroform, precipitated with ethanol and resuspended in the appropriate buffer. The DNA was digested with *Dpn*I or *Mbo*I, extracted with phenol-chloroform and chloroform, precipitated with ethanol and resuspended in formamide-EDTA gel-loading buffer. Samples were electrophoresed on a 12% denaturing polyacrylamide gel, which was subsequently dried and autoradiographed.

The restriction digest showed that the pZTA1 plasmid DNA was completely methylated at the control GATC sites: *Dpn*I cleavage was complete, and no cutting by *Mbo*I was detected. However, the GATC sites in the $(CG)_n$ inserts were not cleaved to completion with *Dpn*I, and the degree of inhibition varied from 16% for pZTA1 grown in DL655, to ~34% for plasmids isolated from the other hosts. While this inhibition of *Dpn*I cleavage may indicate undermethylation, no reciprocal cleavage with *Mbo*I was observed. It is therefore uncertain whether this result may be interpreted as inhibition of Dam methylation, or merely as inhibition of *Dpn*I cleavage.

Conclusion

The results presented in the first part of this chapter show that a $(GC)_n$ insert capable of forming Z-DNA is not stable when cloned in the left arm of λ. This instability may be the consequence of elevated levels of negative supercoiling in transcriptionally active regions. However, eliminating leftward transcription from p_L does not prevent the deletion of $(GC)_n$ inserts. The continuing instability may be due to an interaction of RecA protein with the Z-DNA sequence, or rightward transcription from p_R, proceeding through the $(GC)_n$ insert. A *recA* host proved to be incompatible with the *N nin*5 χ° phage used here; a nut_L phage would be the ideal vector for an investigation of the effect of *recA* mutations on $(GC)_n$ stability in λ. Furthermore, the insertion of the $(GC)_n$ repeat to the right of Q-unresponsive terminators would prevent p_R-initiated transcription causing deletion of Z-DNA.

The second part of this chapter described how a similar $(CG)_n$ repeat was cloned stably in the *Eco*RI site of pBR322. This region of the plasmid is not transcribed, and other inserts capable of Z-DNA formation have been successfully inserted at the same position (Jaworski *et al.*, 1989). The stability of the $(CG)_n$ repeat was unaffected by growth in *E. coli* hosts mutant for *recA* and/or *sbcC*. However, this does not necessarily imply that RecA does not interact with Z-DNA. The $(CG)_n$ insert may undergo the B-Z transition infrequently and transiently, and consequently neither interacts with RecA nor undergoes deletion.

An attempt was made to assay the efficiency of Dam methylation at a GATC site in the $(CG)_n$ insert. Undermethylation of this target may constitute evidence for the formation of Z-DNA by this sequence *in vivo*. Although an inhibition of *Dpn*I cleavage was detected at the GATC site in the $(CG)_n$ insert (and not at control GATC sites) no reciprocal cleavage with *Mbo*I was seen. It is therefore difficult to conclude that the GATC site in the $(CG)_n$ insert is undermethylated *in vivo*. However, it is possible that the inhibition of Dam methylation by Z-DNA formation cannot be detected here as the plasmid is highly methylated and the assay is therefore insensitive (see Introduction to Chapter 3).

In summary, this chapter shows that a $(GC)_n$ repeat cannot be cloned stably in bacteriophage λ. This instability is not exclusively due to transcription initiated at the p_L promoter. Conversely, the insertion of a similar $(CG)_n$ into the *Eco*RI site of pBR322 is possible. The stability of this insert is not affected by *recA* and/or *sbcC* mutations. However, Z-DNA formation *in vivo* could not be detected unequivocally using Dam methylase.

CHAPTER 8

EFFECT OF PALINDROME CENTRAL SEQUENCE ON DAM METHYLATION *IN VIVO*

Introduction

As described in the Introduction to this thesis, the composition of the central sequence of a palindrome profoundly influences the rate of cruciform extrusion (Murchie & Lilley, 1987; Courey & Wang, 1988; Zheng & Sinden, 1988). In particular, the thermal stability of the central ~10 bp of a palindrome correlates inversely with the rate of S-type cruciform extrusion *in vitro* (Murchie & Lilley, 1987). The effect of central sequence changes on undermethylation of the palindrome GATC site *in vivo* was therefore investigated.

Results

Changes to Palindrome Central Sequence

A set of four palindromic oligonucleotides was used to introduce changes to the central sequence of the long DNA palindrome in bacteriophage λ as described previously. The changes are in concerted positions outside the central GATC site, and alter the predicted thermal stability of the DNA.

Four Palindromic Oligonucleotides with Central GATC Sites

The four oligonucleotides shown in Table 8.1 are similar to the sequence described in Chapter 4. Each contains a central target site for Dam methylase. However, the concerted changes to the base pairs immediately outside the GATC sequence create target sites for four different restriction endonucleases; they are shown in Table 8.1. Also shown are the predicted thermal stabilities (T_{MN}) of the central 10 bp of the appropriate palindromes, which were calculated using published parameters (Gotoh & Tagashira, 1981).

Table 8.1

Sequence and predicted thermal stability (T_{MN}) of palindromic oligonucleotide centres.

Oligonucleotide Sequence	Restriction Endonuclease	Thermal Stability (°C)
5' GT**G**GATC**C**ACAGCT 3' 3' TCGACA**C**CTAG**G**TG 5'	*Bam*HI	78.6
5' GT**T**GATC**A**ACAGCT 3' 3' TCGACA**A**CTAG**T**TG 5'	*Bcl*I	71.6
5' GT**A**GATC**T**ACAGCT 3' 3' TCGACA**T**CTAG**A**TG 5'	*Bgl*II	68.5
5' GT**C**GATC**G**ACAGCT 3' 3' TCGACA**G**CTAG**C**TG 5'	*Pvu*I	82.6

Cloning of Oligonucleotides in Palindrome Centre

The oligonucleotides shown in Table 8.1 were cloned in the centre of the 462 bp palindrome in λ as described in Chapter 4. However, λ DRL167 was used as a vector for the palindrome in this instance. DRL167 (Δ*B pal*462 *spi*6 *cI*857 χ^+C153) is a χ^+ derivative of DRL133 (see Chapter 4) and had been constructed by a cross of DRL133 (Δ*B pal*462 *spi*6 *cI*857) with DRL152 (*spi*6 *cI*857 χ^+C153) (Davison & Leach, in the press). The cloning procedure was otherwise identical to that described in Chapter 4.

Selection of λ Phage with Oligonucleotide Insertion in Palindrome

The *in vitro* packaged phage were amplified by one round of plating on JC7623 (*recBC sbcBC*). 26 plaques corresponding to each oligonucleotide insertion were picked into phage buffer and spotted onto bacterial lawns of N2361 (*rec*$^+$) and JC7623. Phage growing on JC7623 but not on N2361 were selected and purified by two further rounds of growth on JC7623. Single plaques were picked for the preparation of phage stocks by plate lysates, again using JC7623 as a host strain.

λ DNA was prepared from the plate lysates and cut with *Sac*I to identify parental phage: none was cleaved. The phage DNA samples were then cleaved with the appropriate restriction endonucleases (the enzymes specific for the oligonucleotide insertions) and electrophoresed on a 0.7% agarose gel. Insertions of the *Bam*HI, *Bcl*I and *Bgl*II palindrome centres were readily discernible by novel restriction digest fragments. However, in the case of the *Pvu*I palindrome centre it proved necessary to cut the λ DNA with *Pvu*I and *Eco*RI, radiolabel with [α^{35}S]dATP, and electrophorese the fragments on a 6% denaturing polyacrylamide gel. The presence of a *Pvu*I site (at bp 26258 of λ) only 390 bp from the palindrome centre precluded the use of agarose gel electrophoresis to identify novel DNA fragments produced by a *Pvu*I restriction digest. The recombinant phage identified by virtue of their new restriction endonuclease target sites were designated as shown in Table 8.2.

Table 8.2

Phage strains constructed by insertion of oligonucleotides at palindrome centre.

Bacteriophage Strain	DRL176	DRL177	DRL178	DRL179
Palindrome Central Site	*Bam*HI	*Bcl*I	*Bgl*II	*Pvu*I

Analysis of *in Vivo* Dam Methylation of DRL176 – 179

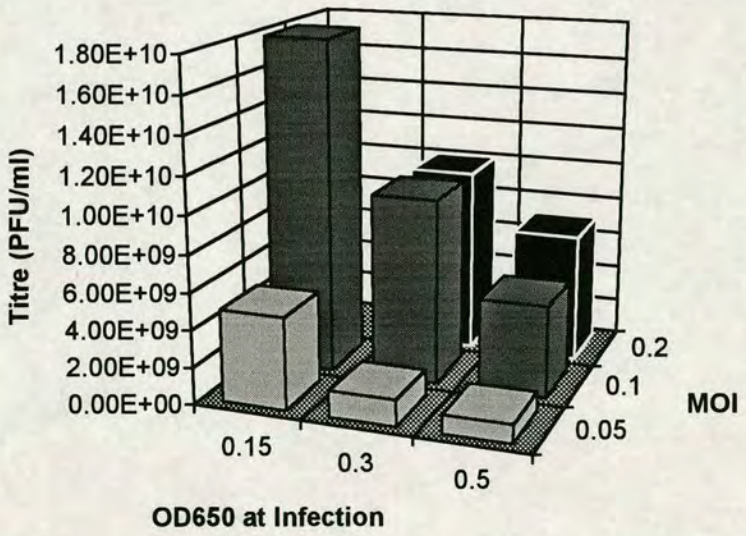
In Vivo Methylation by Growth in N2364 and JC7623 Hosts

DRL176 – 179 were used to infect N2364 (*sbcC dam*$^+$) and JC7623 (*recBC sbcBC dam*$^+$) at a MOI of 0.1, and grown by liquid lysis. After complete cell lysis, phage particles were purified and λ DNA extracted using the large scale method. While the yield of DRL176 – 178 DNA was adequate, very little DRL179 DNA was recovered; DRL179 demonstrated a significant reduction in phage viability, which was particularly apparent in the N2364 host.

The conditions of growth by liquid lysis were therefore altered to maximise the yield of DRL179. The MOI was varied from 0.05 to 0.2, while the cell density of N2364 at the time of infection was varied from OD_{650} 0.15 to 0.5. After complete cell lysis, which took up to 8 hours from the time of infection, the phage suspension was serially diluted and spotted on a lawn of JC7623. The titres are shown in Fig. 8.1.

Figure 8.1

Titre of DRL179 grown in N2364 (see text for details).



The graph in Fig. 8.1 shows that greatest increase in phage yield is brought about by infecting the host cells earlier than the usual OD_{650} of 0.5. Increasing the MOI beyond the usual value of 0.1 does not appear to benefit the titre of DRL179 significantly. N2364 and JC7623 were therefore grown to an OD_{650} of 0.15 and infected with DRL179 at a MOI of 0.1. After complete cell lysis, phage particles were purified and λ DNA extracted using the large scale method.

Analysis of Methylation at GATC Sites in DRL176 – 179

The extent of Dam methylation at the palindrome centres was assayed by *Dpn*I and *Mbo*I cleavage as described in Chapter 4; [α^{35}S]dATP was used to radiolabel the *Eco*RI termini as described in Chapter 5. The appropriate restriction endonucleases (see Tables 8.1 and 8.2) were used in the place of *Bam*HI to cleave the palindrome centre to completion. The samples were electrophoresed as described earlier and autoradiographed to provide the images in Figs. 8.2, 8.5 and 8.10, or exposed to a storage phosphor screen for the quantification of band intensities.

Quantification of Methylation at GATC Sites

The bands in lanes corresponding to *Dpn*I and *Mbo*I cleavage in Figs. 8.2, 8.5 and 8.10 were quantified by PhosphorImager™ and their values are shown in Tables 8.3 – 8.10. The level of Dam methylation at GATC sites *in vivo* and the relative efficiency of *Dpn*I and *Mbo*I cleavage at these sites were calculated as for Table 4.1, and are also shown in Tables 8.3 – 8.10. The latter values are plotted in Figs. 8.3, 8.4, 8.6 – 8.9, 8.11 and 8.12. In each case, the data and graphs are presented with the appropriate autoradiograph. The analysis of Dam methylation at each palindrome centre has been repeated at least once, with very similar results to those shown here.

Fig. 8.10 shows the restriction digest at GATC sites in DRL179. A number of bands can be seen in the lanes 1 and 4, which show cleavage by *Eco*RI and *Pvu*I, that do not correspond to known *Eco*RI-*Pvu*I fragments. These bands (labelled $\Delta 1$ – $\Delta 4$) are particularly apparent in lanes 1 – 3, which shows DRL179 DNA isolated from N2364. They are most probably due to deletions of the palindromic sequence, as their intensity is commensurate with a reduction in the recovery of the palindrome. However, this does not imply that the palindrome in DRL179 is exceptionally unstable. Rather, the reduced viability of DRL179, which is particularly apparent in N2364, confers a significant selective advantage on fast-growing deletion derivatives.

The intensity of these additional bands was quantified and used to correct the quantified intensities of the full-length and half-length palindrome bands in lanes 2, 3, 5 and 6 (*Dpn*I and *Mbo*I cleavage) of Fig. 8.10. For example, the corrected value for the half-length palindrome band is given by:

$$\text{½Pal}_{\text{Corrected}} = \text{½Pal}_{\text{Uncorrected}} \times \frac{(\text{½Pal}_{\text{Pvu}} + \text{Pal}_{\text{Pvu}} + \Delta 1 + \Delta 2 + \Delta 3 + \Delta 4)}{(\text{½Pal}_{\text{Pvu}} + \text{Pal}_{\text{Pvu}})} \qquad (13)$$

By correcting for deletion of the palindrome in this manner, it is possible to compare the intensity of the half palindrome band to the intensities of control bands.

Figure 8.2

Autoradiograph of Restriction Digests at GATC Sites in DRL176

Autoradiograph showing restriction digests of λ DRL176 DNA determining the level of Dam methylation at various GATC sites *in vivo*. Two sets of digests are shown, representing methylation in different *E. coli* hosts. The λ phage had been grown in either N2364 (*sbcC*) or JC7623 (*recBC sbcBC*) and DNA purified as described in the text. The DNA was cleaved with *Eco*RI and radiolabelled with [α^{35}S]dATP as described in the text, before cleavage with *Bam*HI, *Dpn*I or *Mbo*I. The *E. coli* host strain and the appropriate enzymes are indicated above each lane. The full-length and half-length palindrome, and the control fragments (see Fig. 4.3) are shown by the labels on the right of the autoradiograph.

From left to right, the lanes are:

1. λ DRL176 [N2364 (*sbcC*) host] *Eco*RI *Bam*HI
2. λ DRL176 [N2364 (*sbcC*) host] *Eco*RI *Dpn*I
3. λ DRL176 [N2364 (*sbcC*) host] *Eco*RI *Mbo*I
4. λ DRL176 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Bam*HI
5. λ DRL176 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Dpn*I
6. λ DRL176 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Mbo*I

Cleavage by *Bam*HI at the central GGATCC site generates the half-length palindrome; the origin of the doublet band is not known (see Chapter 5). Digestion with *Dpn*I indicates the level of fully methylated DNA, and *Mbo*I the level of fully unmethylated DNA.

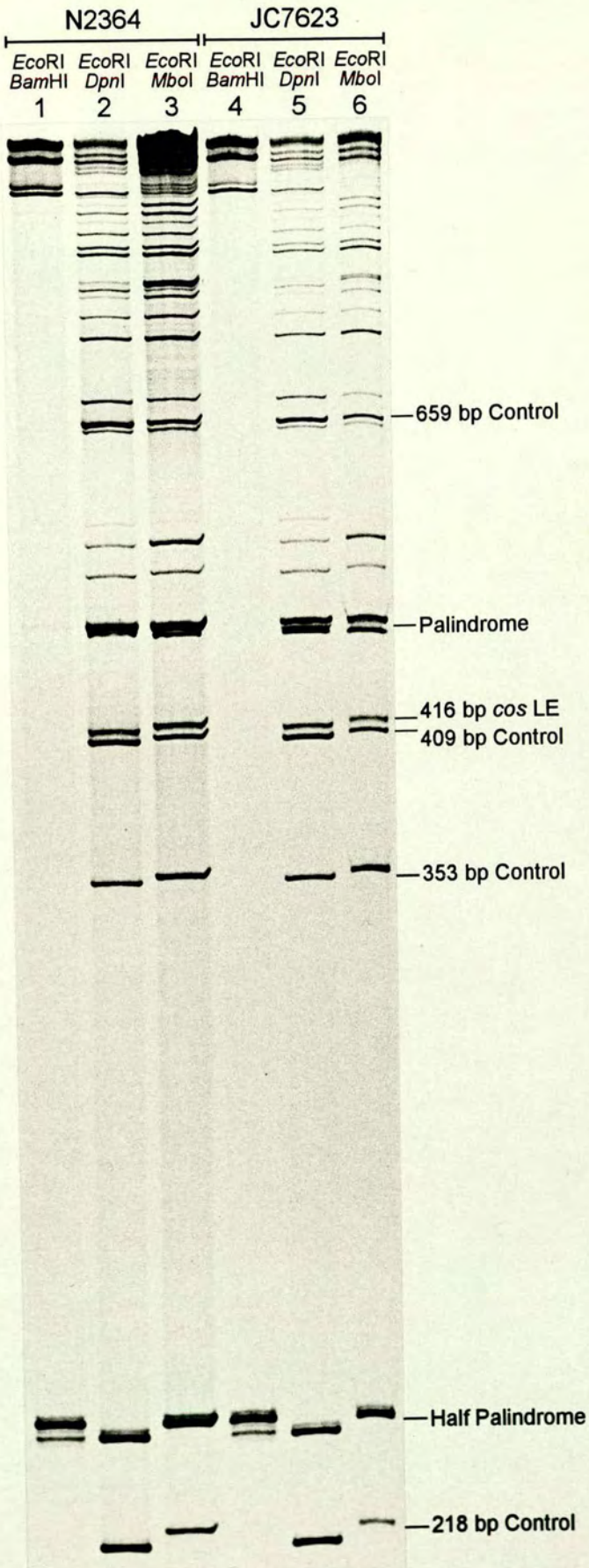

N2364
JC7623
EcoRI BamHI
EcoRI DpnI
EcoRI MboI
EcoRI BamHI
EcoRI DpnI
EcoRI MboI
1
2
3
4
5
6
659 bp Control
Palindrome
416 bp cos LE
409 bp Control
353 bp Control
Half Palindrome
218 bp Control

Table 8.3

Methylation in N2364 (*sbcC*) of GATC Sites in DRL176

The bands shown in lanes 2 and 3 of Fig. 8.2, which correspond to methylation in N2364 (*sbcC*), were quantified using a PhosphorImager™. Their intensities are displayed in Table 8.3 in the rows designated "Data". The quantified values (Uncorrected Data) for the full-length and half-length palindrome were halved (Corrected Data). The % cleavage at the palindrome centre by *Dpn*I or *Mbo*I was calculated using equation (9) and the % methylation at the palindrome centre was determined using the level of methylated DNA as measured by both *Dpn*I and *Mbo*I cleavage; the values from the two enzymes were averaged using equation (10). The average % cleavage at control sites was determined by equation (11) using the sum of the palindrome bands as a standard, and the average % methylation of the control sites was similarly calculated using equation (12). The bottom portion of the table shows the relative efficiency of cleavage by *Dpn*I and *Mbo*I at the palindrome and control GATC sites. For a detailed explanation of these calculations see Table 4.1

Table 8.4

Methylation in JC7623 (*recBC sbcBC*) of GATC Sites in DRL176

The bands shown in lanes 5 and 6 of Fig. 8.2, which correspond to methylation in JC7623 (*recBC sbcBC*), were quantified using a PhosphorImager™. Their intensities are displayed in Table 8.4 in the rows designated "Data". The values derived from these data were calculated as for Table 8.3 (see above).

Table 8.3

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$		*Mbo*I (GATC)
Uncorrected Data	Palindrome	23410		28921
	½ Palindrome	15187		16865
Corrected Data	Palindrome	11705		14461
	½ Palindrome	7594		8433
% Cleavage at Palindrome Centre		39.4		36.8
% Methylation of Palindrome Centre			51.3	
Control Data	659 bp Control	13406		5490
	409 bp Control	10175		3692
	353 bp Control	8858		8974
	218 bp Control	12045		3301
	Average Control	11121		5364
Average % Cleavage at Control Sites		57.6		23.4
Average % Methylation of Control Sites			67.1	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	68.3		157.2
	669 bp Control	120.6		102.3
	409 bp Control	91.5		68.8
	353 bp Control	79.7		167.3
	218 bp Control	108.3		61.5
	Average Control	100		100

Table 8.4

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$		*Mbo*I (GATC)
Uncorrected Data	Palindrome	4353		7955
	½ Palindrome	2504		5596
Corrected Data	Palindrome	2177		3978
	½ Palindrome	1252		2798
% Cleavage at Palindrome Centre		36.5		41.3
% Methylation of Palindrome Centre			47.6	
Control Data	659 bp Control	2360		2447
	409 bp Control	2081		1783
	353 bp Control	1754		3316
	218 bp Control	2327		1465
	Average Control	2131		2253
Average % Cleavage at Control Sites		62.1		33.3
Average % Methylation of Control Sites			64.5	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	58.8		124.2
	669 bp Control	110.8		108.6
	409 bp Control	97.7		79.2
	353 bp Control	82.3		147.2
	218 bp Control	109.2		65.0
	Average Control	100		100

Figure 8.3

Relative Cleavage by *Dpn*I at GATC Sites in DRL176

Graph showing the relative efficiency of cleavage by *Dpn*I at GATC sites in λ DRL176 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨). The values plotted are those shown on the left-hand side of the bottom portions of Tables 8.3 and 8.4. The intensities of the bands shown in Fig. 8.2 are presented relative to the average intensity of the control bands, although the value for the half-length palindrome band was halved to correct for radiolabelling at two *Eco*RI sites. Error bars are also shown, representing the range of relative cleavage by *Dpn*I at GATC sites as calculated from individual restriction digests. Figure 8.3 represents graphically the presence of Dam methylation at various GATC sites in DRL176.

Figure 8.4

Relative Cleavage by *Mbo*I at GATC Sites in DRL176

Graph showing the relative efficiency of cleavage by *Mbo*I at GATC sites in λ DRL176 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨). The values plotted are those shown on the right-hand side of the bottom portions of Tables 8.3 and 8.4. The data are presented as in Fig. 8.3. Error bars are also shown, representing the range of relative cleavage by *Mbo*I at GATC sites as calculated from individual restriction digests. Figure 8.4 represents graphically the absence of Dam methylation at various GATC sites in DRL176.

Figure 8.3

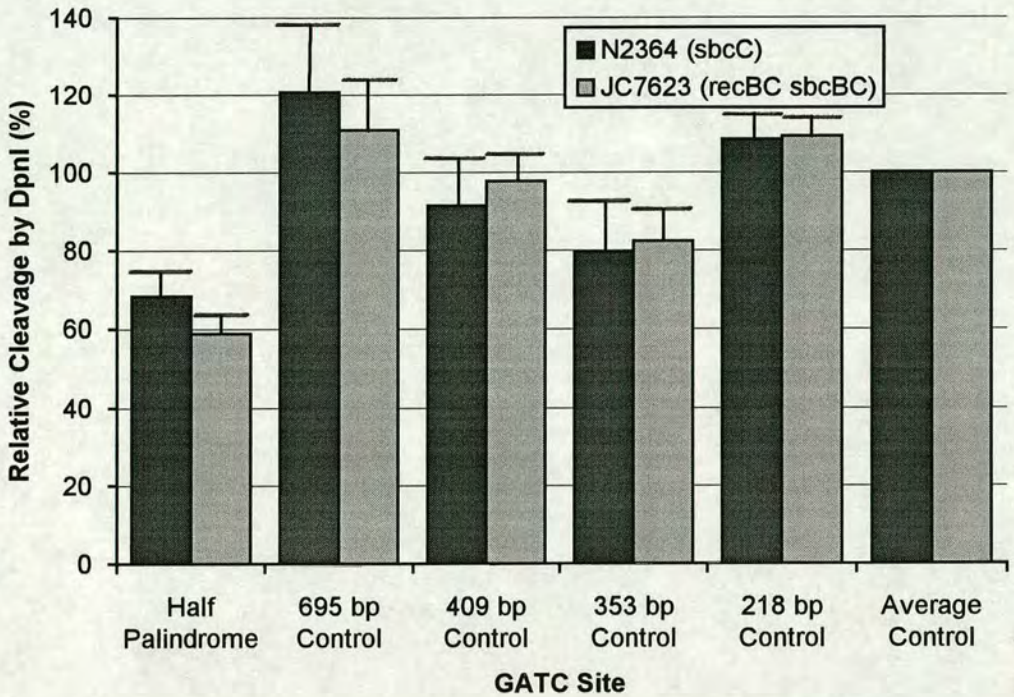


Figure 8.4

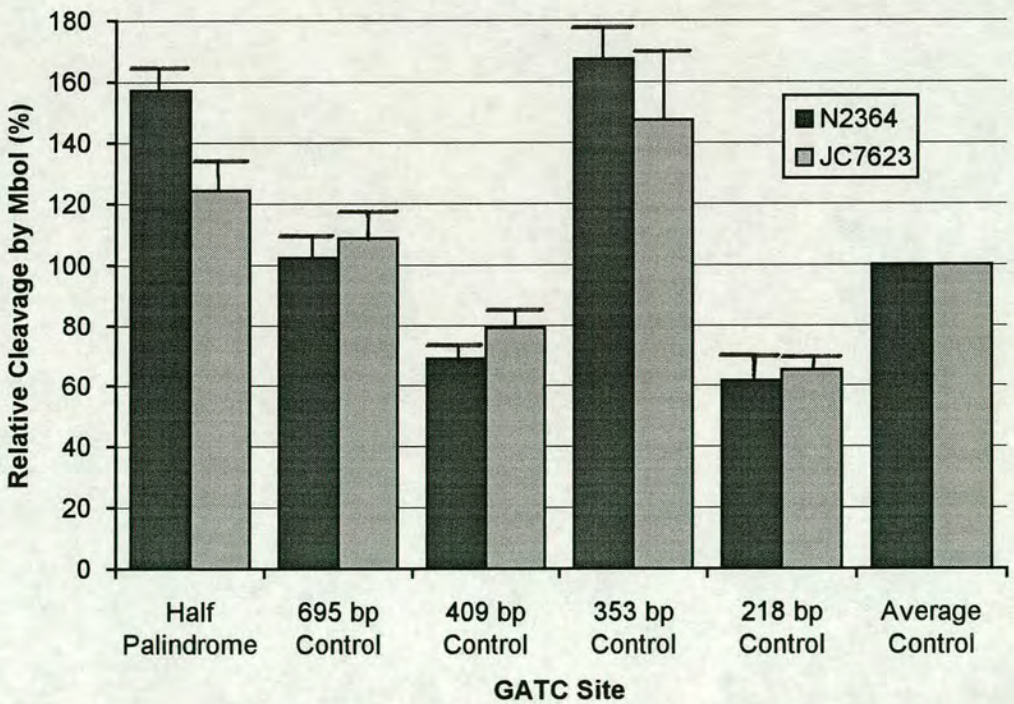


Figure 8.5

Autoradiograph of Restriction Digests at GATC Sites in DRL177 – 178

Autoradiograph showing restriction digests of λ DRL177 and DRL178 DNA determining the level of Dam methylation at various GATC sites *in vivo*. Two sets of digests are shown for each phage, representing methylation in different *E. coli* hosts. The λ phage had been grown in either N2364 (*sbcC*) or JC7623 (*recBC sbcBC*) and DNA purified as described in the text. The DNA was cleaved with *Eco*RI and radiolabelled with [α^{35}S]dATP as described in the text, before cleavage with *Bcl*I (DRL177) or *Bgl*II (DRL178), *Dpn*I or *Mbo*I. The λ strain, the *E. coli* host and the appropriate enzymes are indicated above each lane. The full-length and half-length palindrome, and the control fragments (see Fig. 4.3) are shown by the labels on the right of the autoradiograph.

From left to right, the lanes are:

1. λ DRL177 [N2364 (*sbcC*) host] *Eco*RI *Bcl*I
2. λ DRL177 [N2364 (*sbcC*) host] *Eco*RI *Dpn*I
3. λ DRL177[N2364 (*sbcC*) host] *Eco*RI *Mbo*I
4. λ DRL177 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Bcl*I
5. λ DRL177 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Dpn*I
6. λ DRL177 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Mbo*I
7. λ DRL178 [N2364 (*sbcC*) host] *Eco*RI *Bgl*II
8. λ DRL178 [N2364 (*sbcC*) host] *Eco*RI *Dpn*I
9. λ DRL178[N2364 (*sbcC*) host] *Eco*RI *Mbo*I
10. λ DRL178 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Bgl*II
11. λ DRL178 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Dpn*I
12. λ DRL178 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Mbo*I

Cleavage by *Bcl*I at a central AGATCT site or by *Bgl*II at a central TGATCA site generates the half-length palindrome. However, *Bcl*I is inhibited by Dam methylation (like *Mbo*I) and hence does not cut the palindrome centre to completion. Digestion with *Dpn*I indicates the level of fully methylated DNA, and *Mbo*I the level of fully unmethylated DNA.

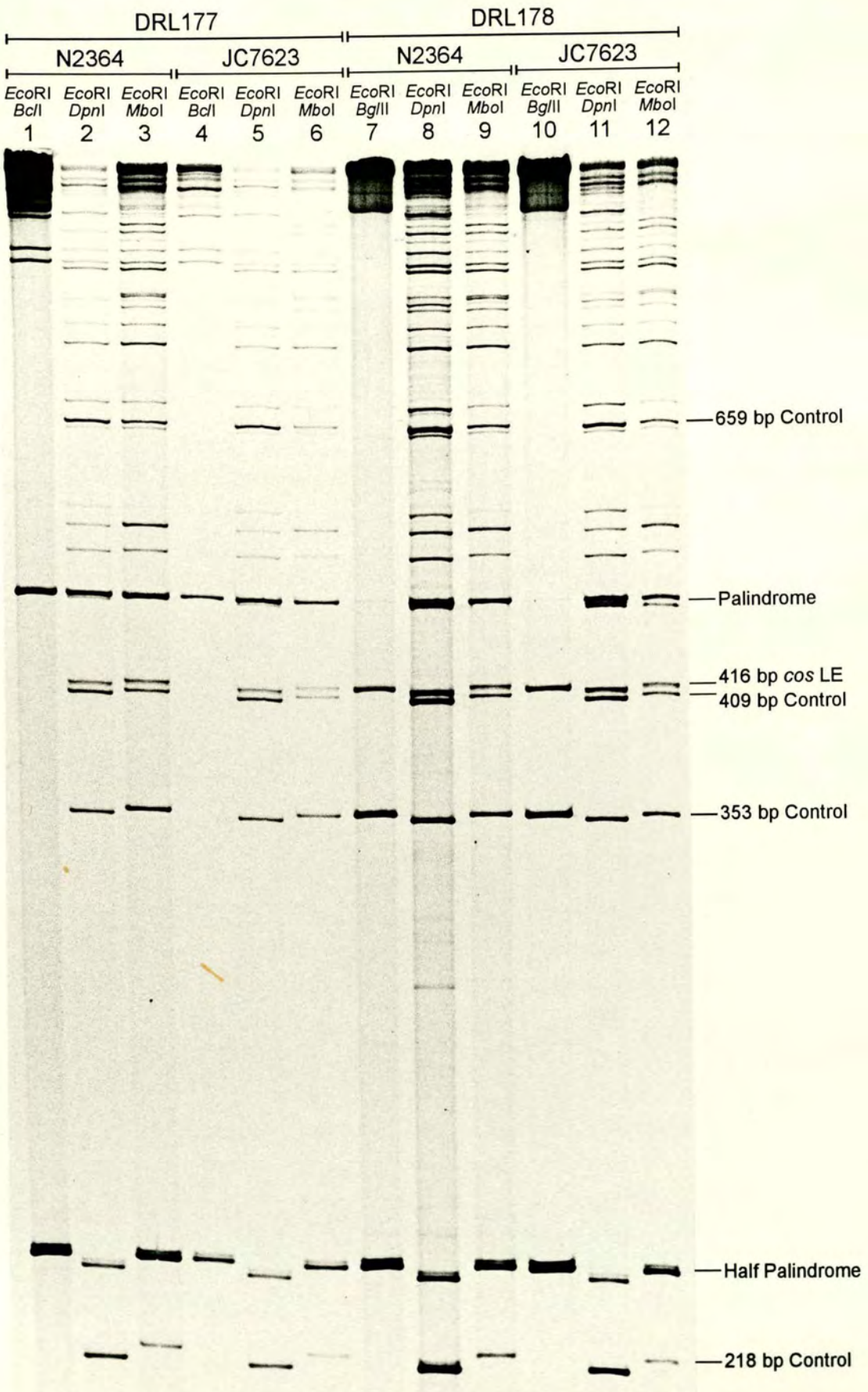

DRL177
DRL178
N2364
JC7623
N2364
JC7623
EcoRI BclI
EcoRI DpnI
EcoRI MboI
EcoRI BclI
EcoRI DpnI
EcoRI MboI
EcoRI BglII
EcoRI DpnI
EcoRI MboI
EcoRI BglII
EcoRI DpnI
EcoRI MboI
1
2
3
4
5
6
7
8
9
10
11
12
659 bp Control
Palindrome
416 bp cos LE
409 bp Control
353 bp Control
Half Palindrome
218 bp Control

Table 8.5

Methylation in N2364 (*sbcC*) of GATC Sites in DRL177

The bands shown in lanes 2 and 3 of Fig. 8.5, which correspond to methylation in N2364 (*sbcC*), were quantified using a PhosphorImager™. Their intensities are displayed in Table 8.5 in the rows designated "Data". The quantified values (Uncorrected Data) for the full-length and half-length palindrome were halved (Corrected Data). The % cleavage at the palindrome centre by *Dpn*I or *Mbo*I was calculated using equation (9) and the % methylation at the palindrome centre was determined using the level of methylated DNA as measured by both *Dpn*I and *Mbo*I cleavage; the values from the two enzymes were averaged using equation (10). The average % cleavage at control sites was determined by equation (11) using the sum of the palindrome bands as a standard, and the average % methylation of the control sites was similarly calculated using equation (12). The bottom portion of the table shows the relative efficiency of cleavage by *Dpn*I and *Mbo*I at the palindrome and control GATC sites. For a detailed explanation of these calculations see Table 4.1

Table 8.6

Methylation in JC7623 (*recBC sbcBC*) of GATC Sites in DRL177

The bands shown in lanes 5 and 6 of Fig. 8.5, which correspond to methylation in JC7623 (*recBC sbcBC*), were quantified using a PhosphorImager™. Their intensities are displayed in Table 8.6 in the rows designated "Data". The values derived from these data were calculated as for Table 8.5 (see above).

Table 8.5

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$		*Mbo*I (GATC)
Uncorrected Data	Palindrome	5372		7591
	½ Palindrome	1466		7332
Corrected Data	Palindrome	2686		3796
	½ Palindrome	733		3666
% Cleavage at Palindrome Centre		21.4		49.1
% Methylation of Palindrome Centre			36.2	
Control Data	659 bp Control	1840		1353
	409 bp Control	1972		1361
	353 bp Control	1311		2581
	218 bp Control	1927		749
	Average Control	1763		1511
Average % Cleavage at Control Sites		51.6		20.3
Average % Methylation of Control Sites			65.7	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	41.6		242.6
	669 bp Control	104.4		89.5
	409 bp Control	111.9		90.1
	353 bp Control	74.4		170.8
	218 bp Control	109.3		49.6
	Average Control	100		100

Table 8.6

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$		*Mbo*I (GATC)
Uncorrected Data	Palindrome	4367		1643
	½ Palindrome	972		2086
Corrected Data	Palindrome	2184		822
	½ Palindrome	486		1043
% Cleavage at Palindrome Centre		18.2		55.9
% Methylation of Palindrome Centre			31.1	
Control Data	659 bp Control	1732		387
	409 bp Control	1557		442
	353 bp Control	1039		711
	218 bp Control	1324		322
	Average Control	1413		466
Average % Cleavage at Control Sites		52.9		25.0
Average % Methylation of Control Sites			64.0	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	34.4		224.1
	669 bp Control	122.6		83.1
	409 bp Control	110.2		95.0
	353 bp Control	73.5		152.7
	218 bp Control	93.7		69.2
	Average Control	100		100

Figure 8.6

Relative Cleavage by *Dpn*I at GATC Sites in DRL177

Graph showing the relative efficiency of cleavage by *Dpn*I at GATC sites in λ DRL177 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨). The values plotted are those shown on the left-hand side of the bottom portions of Tables 8.5 and 8.6. The intensities of the bands shown in Fig. 8.5 are presented relative to the average intensity of the control bands, although the value for the half-length palindrome band was halved to correct for radiolabelling at two *Eco*RI sites. Error bars are also shown, representing the range of relative cleavage by *Dpn*I at GATC sites as calculated from individual restriction digests. Figure 8.6 represents graphically the presence of Dam methylation at various GATC sites in DRL177.

Figure 8.7

Relative Cleavage by *Mbo*I at GATC Sites in DRL177

Graph showing the relative efficiency of cleavage by *Mbo*I at GATC sites in λ DRL177 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨). The values plotted are those shown on the right-hand side of the bottom portions of Tables 8.5 and 8.6. The data are presented as in Fig. 8.6. Error bars are also shown, representing the range of relative cleavage by *Mbo*I at GATC sites as calculated from individual restriction digests. Figure 8.7 represents graphically the absence of Dam methylation at various GATC sites in DRL177.

Figure 8.6

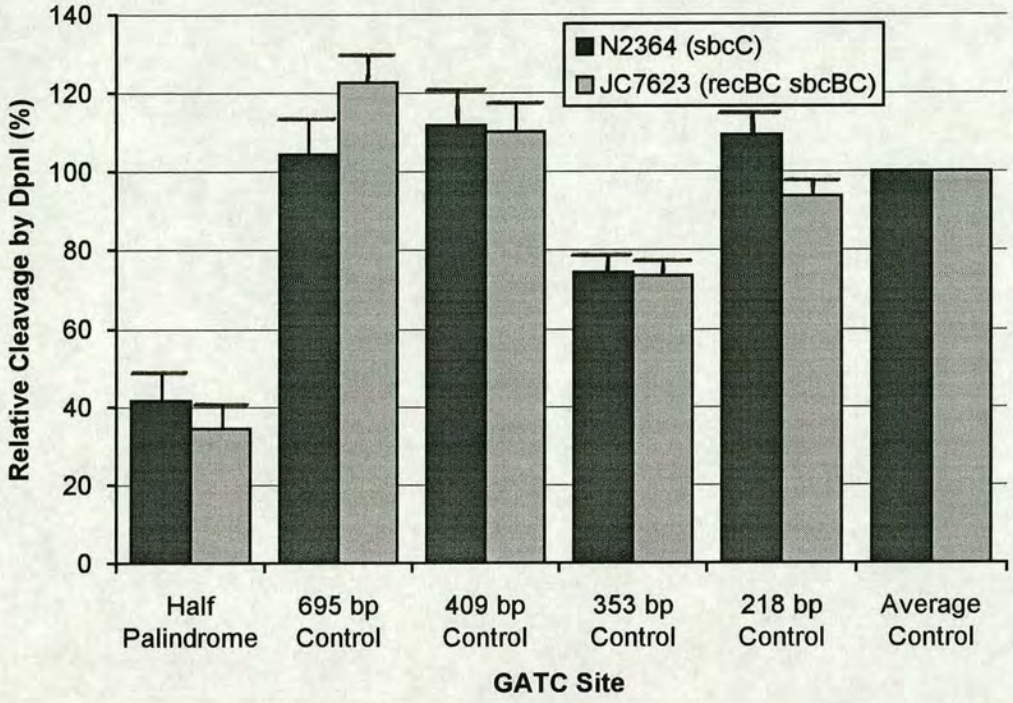


Figure 8.7

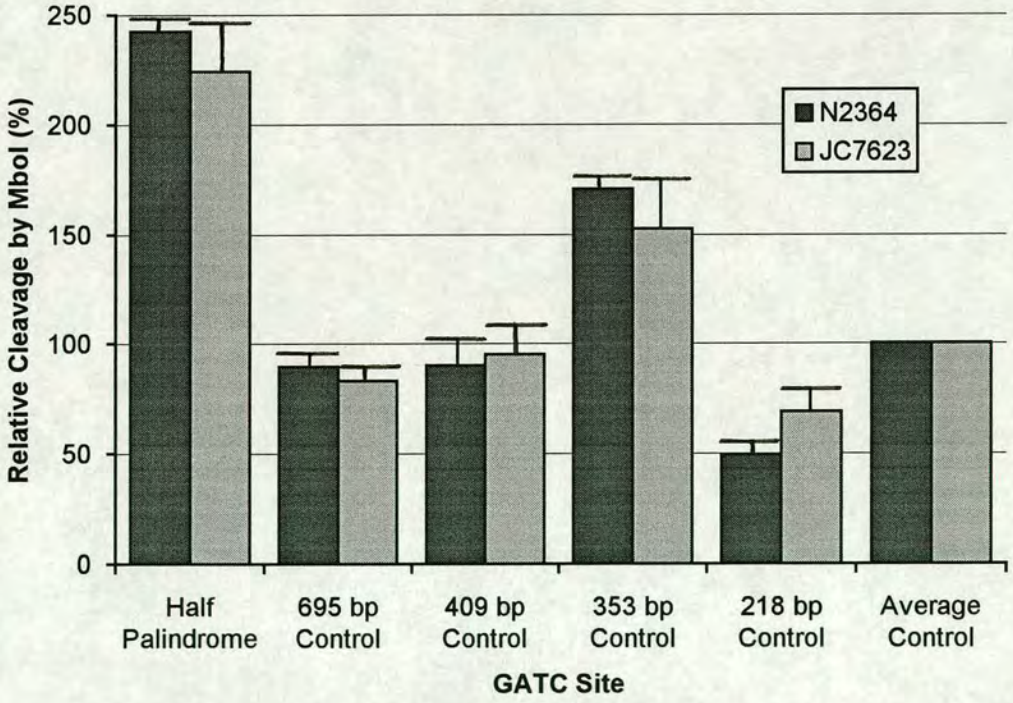


Table 8.7

Methylation in N2364 (*sbcC*) of GATC Sites in DRL178

The bands shown in lanes 8 and 9 of Fig. 8.5, which correspond to methylation in N2364 (*sbcC*), were quantified using a PhosphorImager™. Their intensities are displayed in Table 8.7 in the rows designated "Data". The quantified values (Uncorrected Data) for the full-length and half-length palindrome were halved (Corrected Data). The % cleavage at the palindrome centre by *Dpn*I or *Mbo*I was calculated using equation (9) and the % methylation at the palindrome centre was determined using the level of methylated DNA as measured by both *Dpn*I and *Mbo*I cleavage; the values from the two enzymes were averaged using equation (10). The average % cleavage at control sites was determined by equation (11) using the sum of the palindrome bands as a standard, and the average % methylation of the control sites was similarly calculated using equation (12). The bottom portion of the table shows the relative efficiency of cleavage by *Dpn*I and *Mbo*I at the palindrome and control GATC sites. For a detailed explanation of these calculations see Table 4.1

Table 8.8

Methylation in JC7623 (*recBC sbcBC*) of GATC Sites in DRL178

The bands shown in lanes 11 and 12 of Fig. 8.5, which correspond to methylation in JC7623 (*recBC sbcBC*), were quantified using a PhosphorImager™. Their intensities are displayed in Table 8.8 in the rows designated "Data". The values derived from these data were calculated as for Table 8.7 (see above).

Table 8.7

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$		*Mbo*I (GATC)
Uncorrected Data	Palindrome	27239		5649
	½ Palindrome	5142		9936
Corrected Data	Palindrome	13620		2825
	½ Palindrome	2571		4968
% Cleavage at Palindrome Centre		15.9		63.8
% Methylation of Palindrome Centre			26.1	
Control Data	659 bp Control	12360		1968
	409 bp Control	12888		3107
	353 bp Control	9068		4263
	218 bp Control	14760		1670
	Average Control	12269		2752
Average % Cleavage at Control Sites		75.8		35.3
Average % Methylation of Control Sites			70.2	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	21.0		180.5
	669 bp Control	100.7		71.5
	409 bp Control	105.1		112.9
	353 bp Control	73.9		154.9
	218 bp Control	120.3		60.7
	Average Control	100		100

Table 8.8

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$		*Mbo*I (GATC)
Uncorrected Data	Palindrome	15361		3512
	½ Palindrome	2083		6448
Corrected Data	Palindrome	7681		1756
	½ Palindrome	1042		3224
% Cleavage at Palindrome Centre		11.9		64.7
% Methylation of Palindrome Centre			23.6	
Control Data	659 bp Control	5566		1190
	409 bp Control	5682		1542
	353 bp Control	3923		2738
	218 bp Control	5705		993
	Average Control	5219		1616
Average % Cleavage at Control Sites		59.8		32.4
Average % Methylation of Control Sites			63.7	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	20.0		199.5
	669 bp Control	106.7		73.7
	409 bp Control	108.9		95.4
	353 bp Control	75.2		169.5
	218 bp Control	109.3		61.5
	Average Control	100		100

Figure 8.8

Relative Cleavage by *Dpn*I at GATC Sites in DRL178

Graph showing the relative efficiency of cleavage by *Dpn*I at GATC sites in λ DRL178 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨). The values plotted are those shown on the left-hand side of the bottom portions of Tables 8.7 and 8.8. The intensities of the bands shown in Fig. 8.5 are presented relative to the average intensity of the control bands, although the value for the half-length palindrome band was halved to correct for radiolabelling at two *Eco*RI sites. Error bars are also shown, representing the range of relative cleavage by *Dpn*I at GATC sites as calculated from individual restriction digests. Error bars are also shown, representing the range of relative cleavage by *Dpn*I at GATC sites as calculated from individual restriction digests. Figure 8.8 represents graphically the presence of Dam methylation at various GATC sites in DRL178.

Figure 8.9

Relative Cleavage by *Mbo*I at GATC Sites in DRL178

Graph showing the relative efficiency of cleavage by *Mbo*I at GATC sites in λ DRL178 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨). The values plotted are those shown on the right-hand side of the bottom portions of Tables 8.7 and 8.8. The data are presented as in Fig. 8.8. Error bars are also shown, representing the range of relative cleavage by *Mbo*I at GATC sites as calculated from individual restriction digests. Error bars are also shown, representing the range of relative cleavage by *Mbo*I at GATC sites as calculated from individual restriction digests. Figure 8.9 represents graphically the absence of Dam methylation at various GATC sites in DRL178.

Figure 8.8

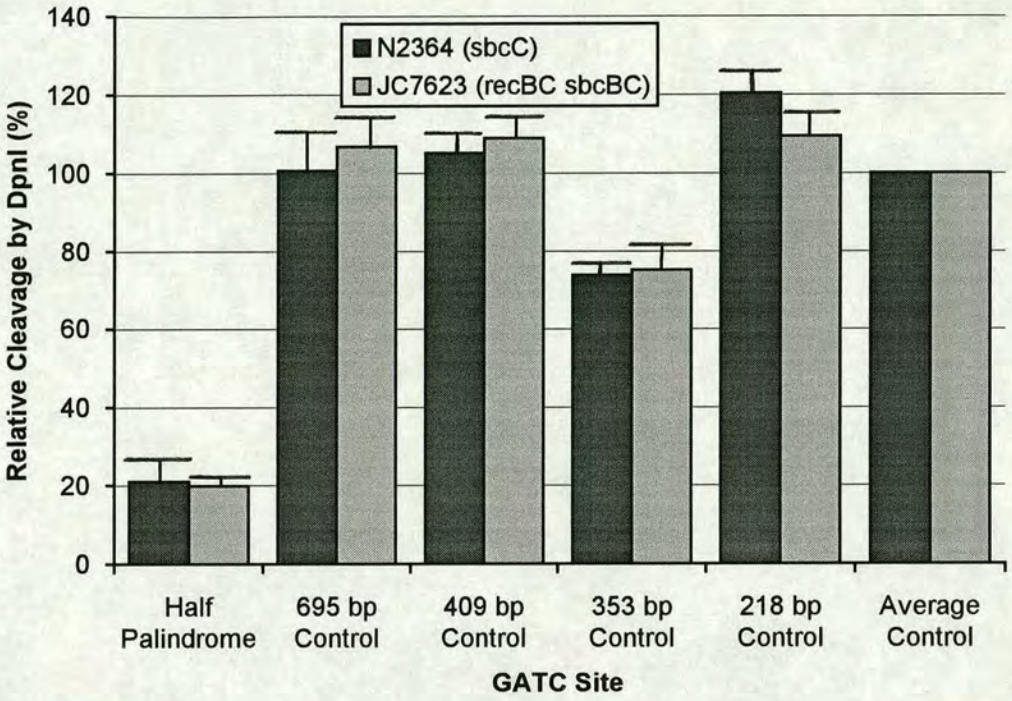


Figure 8.9

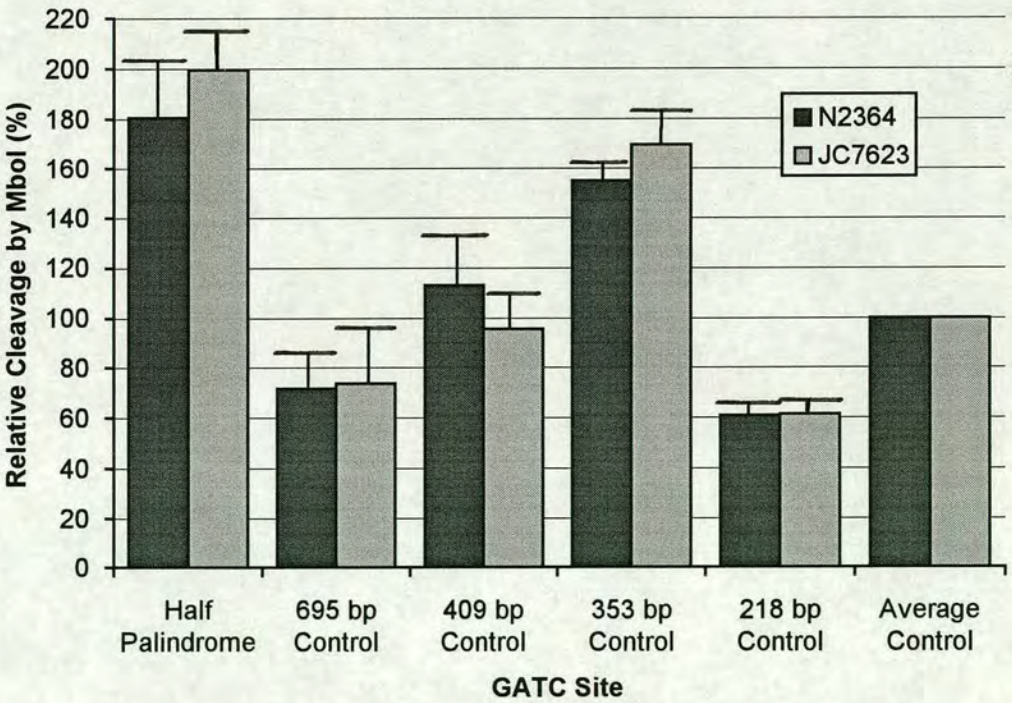


Figure 8.10

Autoradiograph of Restriction Digests at GATC Sites in DRL179

Autoradiograph showing restriction digests of λ DRL179 DNA determining the level of Dam methylation at various GATC sites *in vivo*. Two sets of digests are shown, representing methylation in different *E. coli* hosts. The λ phage had been grown in either N2364 (*sbcC*) or JC7623 (*recBC sbcBC*) and DNA purified as described in the text. The DNA was cleaved with *Eco*RI and radiolabelled with [α^{35}S]dATP as described in the text, before cleavage with *Pvu*I, *Dpn*I or *Mbo*I. The *E. coli* host strain and the appropriate enzymes are indicated above each lane. The full-length and half-length palindrome, and the control fragments (see Fig. 4.3) are shown by the labels on the right of the autoradiograph. Other bands can be seen that are the result of deletions of the palindromic sequence; they are indicated by the labels (Δ1 – Δ4) on the left of the autoradiograph.

From left to right, the lanes are:

1. λ DRL179 [N2364 (*sbcC*) host] *Eco*RI *Pvu*I
2. λ DRL179 [N2364 (*sbcC*) host] *Eco*RI *Dpn*I
3. λ DRL179 [N2364 (*sbcC*) host] *Eco*RI *Mbo*I
4. λ DRL179 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Pvu*I
5. λ DRL179 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Dpn*I
6. λ DRL179 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Mbo*I

Cleavage by *Pvu*I at the central CGATCG site generates the half-length palindrome. Digestion with *Dpn*I indicates the level of fully methylated DNA, and *Mbo*I the level of fully unmethylated DNA.

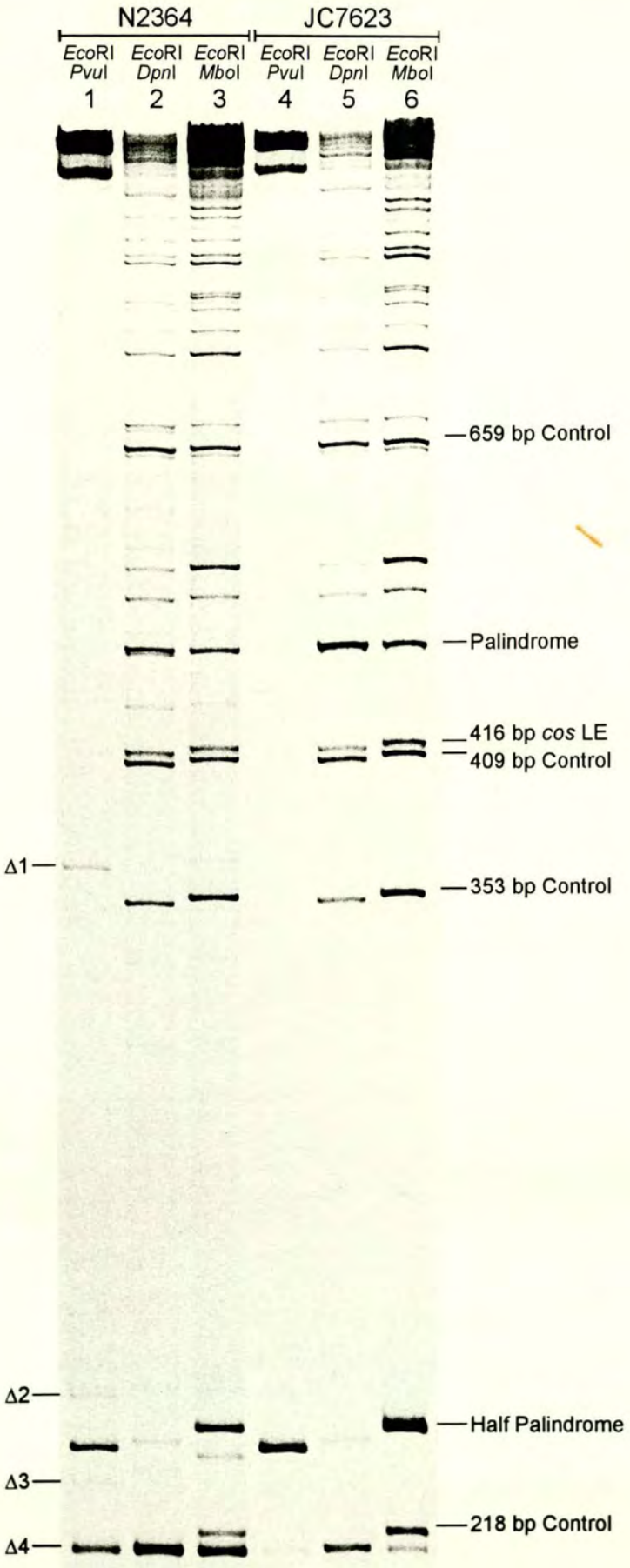

N2364
JC7623
EcoRI PvuI
EcoRI DpnI
EcoRI MboI
EcoRI PvuI
EcoRI DpnI
EcoRI MboI
1
2
3
4
5
6
659 bp Control
Palindrome
416 bp cos LE
409 bp Control
353 bp Control
Half Palindrome
218 bp Control
Δ1
Δ2
Δ3
Δ4

Table 8.9

Methylation in N2364 (*sbcC*) of GATC Sites in DRL179

The bands shown in lanes 2 and 3 of Fig. 8.10, which correspond to methylation in N2364 (*sbcC*), were quantified using a PhosphorImager™. Their intensities are displayed in Table 8.9 in the rows designated "Data". The sum of intensities of the palindrome deletion products (Δ1 – Δ4) in lane 1 of Fig. 8.10 is also shown, and was used to correct the quantified values (Uncorrected Data) for the full-length and half-length palindrome using equation (13) (see earlier parts of this chapter for an explanation).

$$\text{½Pal}_{\text{Corrected}} = \text{½Pal}_{\text{Uncorrected}} \times \frac{(\text{½Pal}_{\text{Pvu}} + \text{Pal}_{\text{Pvu}} + \Delta 1 + \Delta 2 + \Delta 3 + \Delta 4)}{(\text{½Pal}_{\text{Pvu}} + \text{Pal}_{\text{Pvu}})} \qquad (13)$$

The full-length and half-length palindrome intensities were also halved to correct for radiolabelling at the two *Eco*RI sites. Their final values are designated "Corrected Data".

The % cleavage at the palindrome centre by *Dpn*I or *Mbo*I was calculated using equation (9) and the % methylation at the palindrome centre was determined using the level of methylated DNA as measured by both *Dpn*I and *Mbo*I cleavage; the values from the two enzymes were averaged using equation (10). The average % cleavage at control sites was determined by equation (11) using the sum of the palindrome bands as a standard, and the average % methylation of the control sites was similarly calculated using equation (12). The bottom portion of the table shows the relative efficiency of cleavage by *Dpn*I and *Mbo*I at the palindrome and control GATC sites. For a detailed explanation of these calculations see Table 4.1

Table 8.10

Methylation in JC7623 (*recBC sbcBC*) of GATC Sites in DRL179

The bands shown in lanes 5 and 6 of Fig. 8.10, which correspond to methylation in JC7623 (*recBC sbcBC*), were quantified using a PhosphorImager™. Their intensities are displayed in Table 8.10 in the rows designated "Data". The sum of intensities of the palindrome deletion products was quantified from the bands (Δ1 – Δ4) shown in lane 4 of Fig. 8.10. The values derived from these data were calculated as for Table 8.9 (see above).

Table 8.9

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$	*Mbo*I (GATC)
Uncorrected Data	Palindrome	7415	6197
	½ Palindrome	1157	12501
Sum of Palindrome Deletion Products		10456	
Corrected Data	Palindrome	8168	6827
	½ Palindrome	1275	13771
% Cleavage at Palindrome Centre		13.5	66.9
% Methylation of Palindrome Centre		23.3	
Control Data	659 bp Control	8906	6146
	409 bp Control	6834	4775
	353 bp Control	5684	9622
	218 bp Control	8203	4596
	Average Control	7407	6285
Average % Cleavage at Control Sites		78.4	30.5
Average % Methylation of Control Sites		74.0	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	17.2	219.1
	659 bp Control	120.2	97.8
	409 bp Control	92.3	76.0
	353 bp Control	76.7	153.1
	218 bp Control	110.8	73.1
	Average Control	100	100

Table 8.10

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$	*Mbo*I (GATC)
Uncorrected Data	Palindrome	18144	11020
	½ Palindrome	1250	59265
Sum of Palindrome Deletion Products		1680	
Corrected Data	Palindrome	9839	5976
	½ Palindrome	678	32139
% Cleavage at Palindrome Centre		6.5	84.3
% Methylation of Palindrome Centre		11.1	
Control Data	659 bp Control	6691	10389
	409 bp Control	5216	9287
	353 bp Control	3831	16748
	218 bp Control	6301	8938
	Average Control	5510	11341
Average % Cleavage at Control Sites		52.4	29.8
Average % Methylation of Control Sites		61.3	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	12.3	283.4
	659 bp Control	121.4	91.6
	409 bp Control	94.7	81.9
	353 bp Control	69.5	147.7
	218 bp Control	114.4	78.8
	Average Control	100	100

Figure 8.11

Relative Cleavage by *Dpn*I at GATC Sites in DRL179

Graph showing the relative efficiency of cleavage by *Dpn*I at GATC sites in λ DRL179 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨). The values plotted are those shown on the left-hand side of the bottom portions of Tables 8.9 and 8.10. The intensities of the bands shown in Fig. 8.10 are presented relative to the average intensity of the control bands, although the value for the half-length palindrome band was halved on account of radiolabelling at two *Eco*RI sites, and corrected for palindrome deletion by equation (13). Error bars are also shown, representing the range of relative cleavage by *Dpn*I at GATC sites as calculated from individual restriction digests. Figure 8.11 represents graphically the presence of Dam methylation at various GATC sites in DRL179.

Figure 8.12

Relative Cleavage by *Mbo*I at GATC Sites in DRL179

Graph showing the relative efficiency of cleavage by *Mbo*I at GATC sites in λ DRL179 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨). The values plotted are those shown on the right-hand side of the bottom portions of Tables 8.9 and 8.10. The data are presented as in Fig. 8.11. Error bars are also shown, representing the range of relative cleavage by *Mbo*I at GATC sites as calculated from individual restriction digests. Figure 8.12 represents graphically the absence of Dam methylation at various GATC sites in DRL179.

Figure 8.11

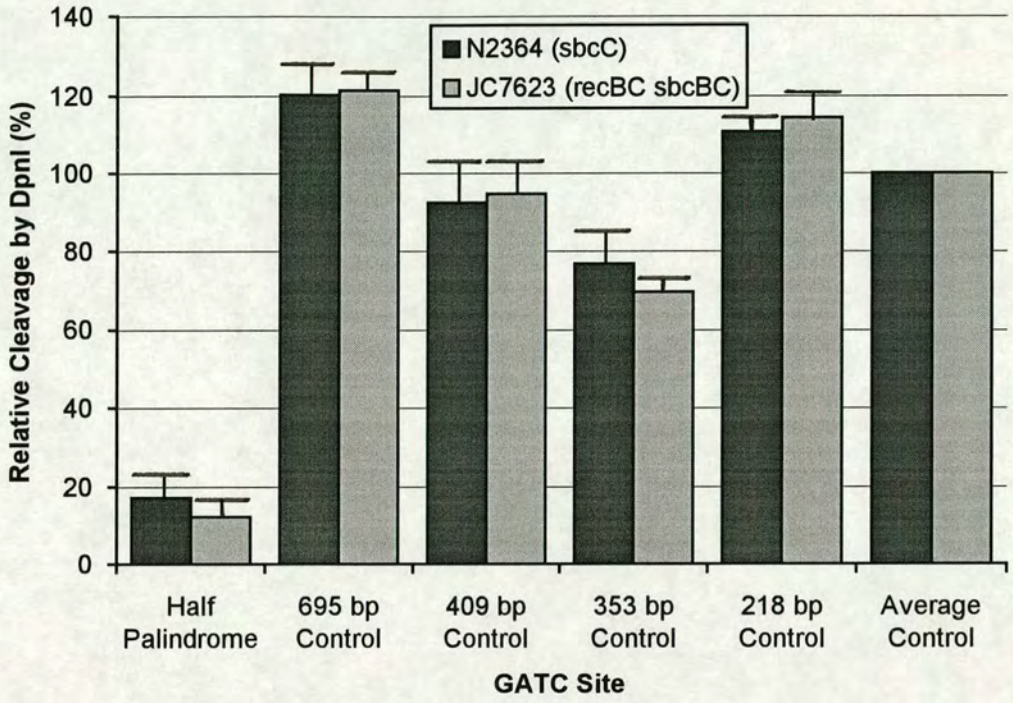
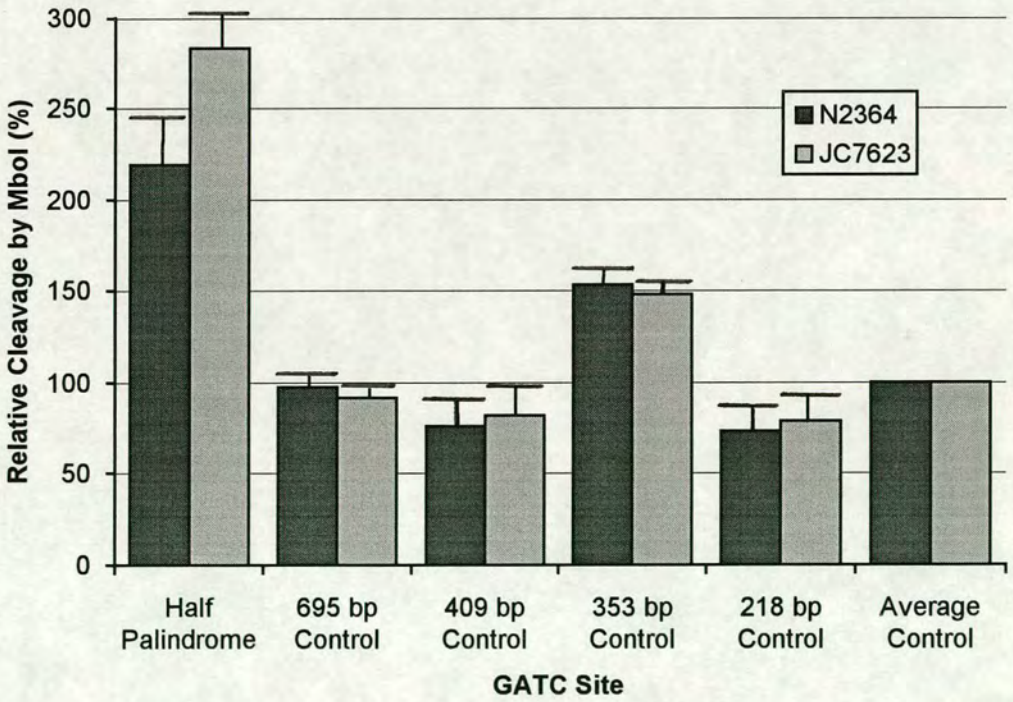


Figure 8.12



Analysis of Methylation Inhibition by Kinetic Modelling

In a similar study using Dam methylase to study protein-DNA interactions *in vivo*, Wang & Church (1992) propose that:

> "The degree of methylation protection can be correlated with that of protein binding through kinetic modelling because protein factors need to bind DNA targets persistently to prevent methylation, whereas methylases need only a brief contact with targets to succeed in methylation. If one assumes that the methylase acts on target sites in a genome at random, then the fraction of targets methylated per unit time is constant, and the kinetic interaction between methylases and protecting factors is characterised by Poisson first-order decay:
>
> $$U = 2^{-[(1-P)G/T]} \qquad (14)$$
>
> where P is the fraction of a cell generation with protein factors that bind to the target sequence, U the unmethylated fraction of the target site at steady state, T the half-life of methylase action, and G the cell generation time."

This form of kinetic modelling can also be used to correlate the undermethylation observed at the palindrome centre with the proportion of time (in *E. coli*) that this GATC site is inaccessible to Dam methylase. This would indicate the proportion of time that a cruciform is extruded, if it is a cruciform structure that is inhibiting Dam methylation at the palindrome centre.

In the kinetic modelling of Wang & Church (1992), the half-life of methylase action (T) was estimated as 4 minutes (Lyons & Schendel, 1984; Campbell & Kleckner, 1990). However, this estimate will probably be inaccurate under unusual conditions such as lytic infection by bacteriophage λ. Furthermore, in the case of the work described in this thesis, G (the generation time) refers to the doubling time of λ and not *E. coli*. If the average burst size of λ is 100 and the time taken for one round of lytic growth is 50 minutes, then the doubling time can be estimated as:

$$\frac{50}{\log_2(100)} = 7\,\text{min}\,31\,\text{sec} \qquad (15)$$

However, this time will also be inaccurate: not all λ DNA molecules are packaged into phage heads, and consequently the burst size of 100 is an underestimate. Moreover, the reduced viability of phage carrying long palindromes (seen most clearly with DRL179) will affect the burst size to varying degrees.

It is possible to circumvent the need to estimate the λ doubling time or the half-time of methylase action. The results presented in Chapter 6 permit the assumption that there is no protection (P) at the control GATC sites, as the relative efficiencies of *in vitro* and *in vivo* Dam methylation at these sites are nearly identical.

If $P = 0$ at control GATC targets, equation (14) can be simplified:

$$U_{Control} = 2^{-[G/T]} \tag{16}$$

Equation (16) may now be rearranged to give an expression for $[G/T]$:

$$[G/T] = \left(-\log_2 U_{Control}\right) \tag{17}$$

Similarly, equation (14) may be rearranged to give an expression for P:

$$P = 1 + \frac{\log_2 U_{Palindrome}}{[G/T]} \tag{18}$$

Equation (17) can now be substituted for the term $[G/T]$ in equation (18) to give:

$$P = 1 - \frac{\log_2 U_{Palindrome}}{\log_2 U_{Control}} \tag{19}$$

$U_{Palindrome}$ can be calculated by averaging the fraction of unmethylated DNA as detected by *Dpn*I and *Mbo*I cleavage:

$$U_{Palindrome} = \frac{\left(1 - \frac{½\text{Pal}_{\text{Dpn}}}{½\text{Pal}_{\text{Dpn}} + \text{Pal}_{\text{Dpn}}}\right) + \left(\frac{½\text{Pal}_{\text{Mbo}}}{½\text{Pal}_{\text{Mbo}} + \text{Pal}_{\text{Mbo}}}\right)}{2} \tag{20}$$

This equation is related to the expression (10) for $\%\text{Methylation}_{\text{Pal}}$ (see Table 4.1). Similarly, $U_{Control}$ can be calculated in an analogous manner to $\%\text{Methylation}_{\text{Control}}$ (12) by using the sum of the palindrome band intensities as a standard:

$$U_{Control} = \frac{\left(1 - \frac{\text{AverageControl}_{\text{Dpn}}}{(½\text{Pal}_{\text{Dpn}} + \text{Pal}_{\text{Dpn}})/2}\right) + \left(\frac{\text{AverageControl}_{\text{Mbo}}}{(½\text{Pal}_{\text{Mbo}} + \text{Pal}_{\text{Mbo}})/2}\right)}{2} \tag{21}$$

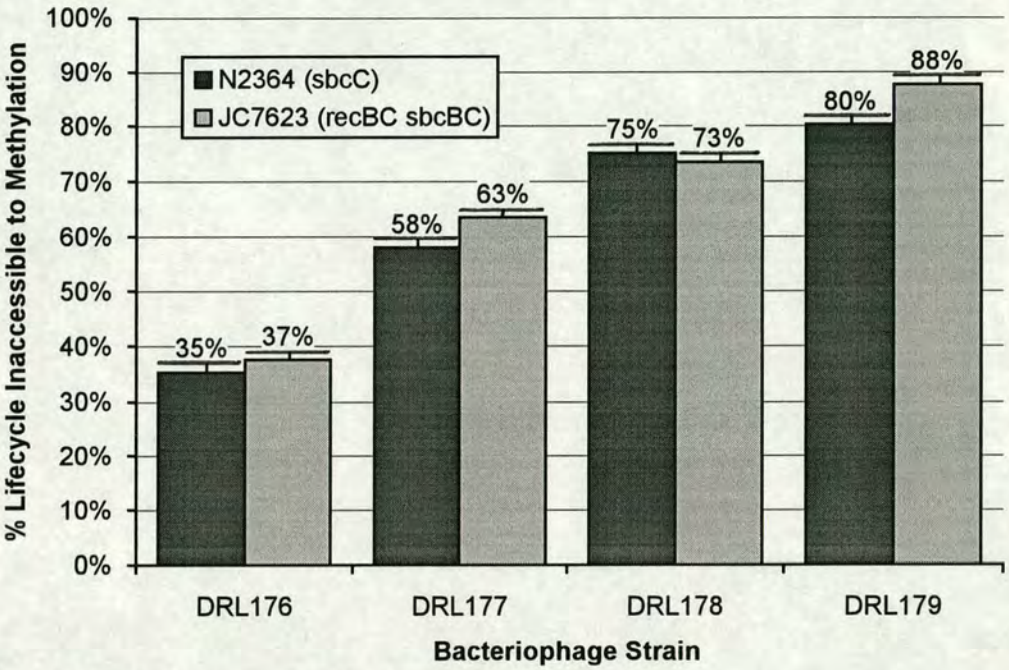
Equations (20) and (21) can now be substituted into (19) to give an expression for the fraction of the λ lifecycle that the palindrome centre is inaccessible to methylation:

$$P = 1 - \left(\frac{\log_2 \left(\frac{\left(1 - \frac{½\text{Pal}_{\text{Dpn}}}{½\text{Pal}_{\text{Dpn}} + \text{Pal}_{\text{Dpn}}}\right) + \left(\frac{½\text{Pal}_{\text{Mbo}}}{½\text{Pal}_{\text{Mbo}} + \text{Pal}_{\text{Mbo}}}\right)}{2} \right)}{\log_2 \left(\frac{\left(1 - \frac{\text{AverageControl}_{\text{Dpn}}}{(½\text{Pal}_{\text{Dpn}} + \text{Pal}_{\text{Dpn}})/2}\right) + \left(\frac{\text{AverageControl}_{\text{Mbo}}}{(½\text{Pal}_{\text{Mbo}} + \text{Pal}_{\text{Mbo}})/2}\right)}{2} \right)} \right) \tag{22}$$

The proportion of the lytic lifecycle of DRL176 – 179 that the palindrome centre is inaccessible to methylation by Dam methylase was calculated using equation (22), and the results are shown in Fig. 8.13.

Figure 8.13

Fraction of λ lytic lifecycle that the palindrome centre is inaccessible to methylation.



This analysis of methylation inhibition using kinetic modelling does not make the assumptions previously required (Wang & Church, 1992) concerning the rate of methylation by Dam methylase and demethylation by replication. It therefore draws on the major strengths of the assay presented here, being internally controlled and quantitative. Furthermore, it expresses the somewhat abstract methylation data in tangible terms.

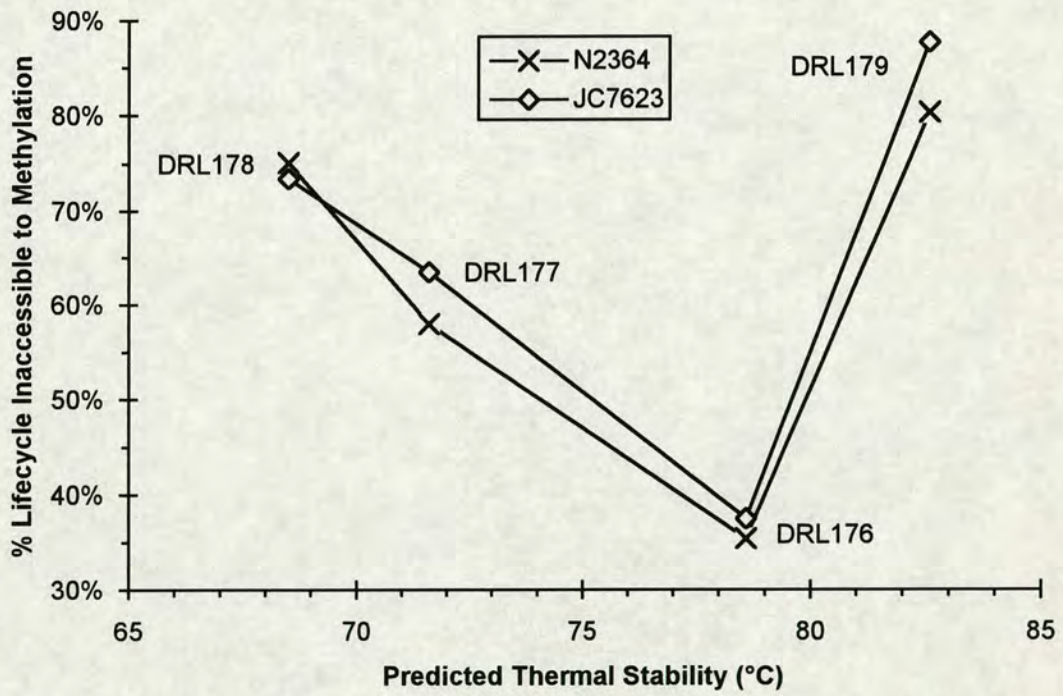
Conclusion

The results presented in this chapter, which are best summarised by Fig. 8.13, demonstrate that the base composition of the palindrome central sequences in DRL176 – 179 profoundly affects methylation at those sites *in vivo*. As these palindrome centres all display significant undermethylation in spite of sequence alterations, it is very unlikely that the serendipitous formation of a protein binding site is responsible for the methylation inhibition. It is also unlikely that the undermethylation is due to an effect of sequence context on Dam methylase activity; the three new centres experience even greater methylation inhibition than the *Bam*HI target previously studied. Furthermore, each of these three new centres shows significantly less methylation than any of the control GATC targets.

If the methylation inhibition at these palindrome centres is a consequence of cruciform extrusion *in vivo*, then the degree of undermethylation might be expected to correlate inversely with the thermal stability of the central ~10 bp of these palindromes. Such a correlation is predicted by *in vitro* studies of S-type cruciform extrusion (Murchie & Lilley, 1987). The relationship between the predicted thermal stability (T_{MN}) (Table 8.1) and the fraction of λ lytic lifecycle that the palindrome centre is inaccessible to methylation (Fig. 8.13) is presented below in Fig. 8.14.

Figure 8.14

Fraction of λ lytic lifecycle that the palindrome centre is inaccessible to methylation vs. predicted thermal stability (T_{MN}) of the central 10 bp of the palindrome.



Although a linear inverse relationship between the T_{MN} of the central 10 bp and the methylation inhibition at the palindrome centre can been seen for DRL176 – 178, this is not the case for DRL179, which shows the greatest undermethylation *in vivo* despite the highest T_{MN} value of 82.6°C. Four possible explanations are:

1. cruciform extrusion does not occur *in vivo*, and the undermethylation is the result of other mechanisms of methylation inhibition;
2. cruciform extrusion occurs *in vivo*, but the undermethylation is also affected by other mechanisms of methylation inhibition;

3. cruciform extrusion occurs *in vivo*, but not by the centre-dependent S-type pathway (see Fig. 1.10);
4. cruciform extrusion by the S-type pathway occurs *in vivo*, but the palindrome centre in DRL179 behaves anomalously.

The alternative mechanism of methylation inhibition invoked in the first and second proposals may be a sequence context effect on Dam methylase, or binding of a protein that recognises long DNA palindromes irrespective of their central sequence. However, these exist only as formal possibilities. The best alternative to cruciform extrusion as a mechanism of methylation inhibition is the formation of a hairpin structure by single-stranded DNA (see Fig. 1.2). This may be facilitated by processes involving DNA unwinding such as transcription. However, it is hard to envisage a dominant influence of the T_{MN} of the palindrome centre on the rate of hairpin formation by single-stranded DNA. Furthermore, it is difficult to account for the high levels of methylation inhibition in terms of hairpin formation by single-stranded DNA.

The third and fourth possibilities are interrelated and cannot be distinguished by the methylation data alone. However, evidence suggesting that the palindrome centre in DRL179 behaves anomalously has emerged from theoretical studies. The predicted free energies (ΔG) of hairpin formation by palindromic oligonucleotides of 10 bp were calculated for the palindrome centres of DRL176 – 179 using the FOLD program (Zuker & Stiegler, 1981) from the University of Wisconsin Genetics Computer Group Sequence Analysis Software Package v7.3. FOLD finds a secondary structure of minimum free energy for an RNA molecule and predicts its energy of formation. The values shown in Table 8.11 therefore refer to RNA hairpins rather than the DNA hairpins shown. They do, however, represent best indication of potential DNA hairpin stabilities, since analogous programs determining DNA secondary structure are not available.

Table 8.11

Predicted free energies of hairpin formation by 10 base oligonucleotides from palindrome centres as calculated by FOLD program (see text for details).

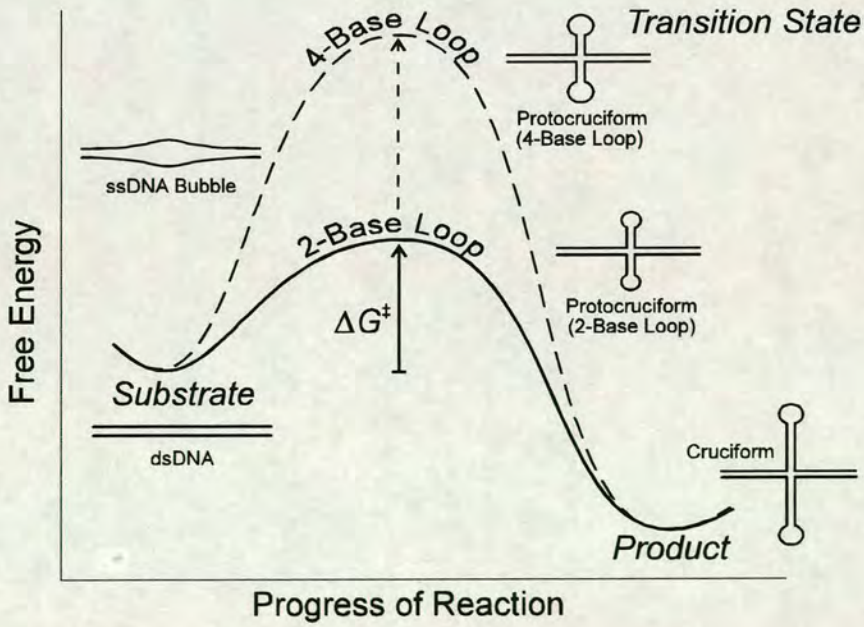
λ Strain	DRL176	DRL177	DRL178	DRL179
Hairpin	G 5'~GTG A 3'~CAC T C	TG 5'~GT A 3'~CA T AC	AG 5'~GT A 3'~CA T TC	G 5'~GTC A 3'~CAG T C
Energy (kcal mol^{-1})	2.0	2.2	2.2	1.5

The values presented in Table 8.11 show that a hairpin formed by the palindrome centre in DRL179 is significantly more stable than the other three molecules. This is due to high stability of base stacking interactions in the 5' TC 3' doublet immediately preceding the loop of DRL179 (Gotoh & Tagashira, 1981).

A palindromic oligonucleotide with a CGATCG central sequence identical to that in DRL179 (Table 8.1) exhibits a high thermal stability consistent with the formation of two-base loops *in vitro* (Xodo *et al.*, 1988). The unusually high stability of the hairpin composed of the palindrome centre of DRL179 (Table 8.11) may similarly facilitate the formation of two-base loops. The extrusion *in vivo* of a cruciform structure with two-membered loops would account for the anomalous behaviour of DRL179. As discussed in the Introduction, the high thermal stability of two-residue loops reduces the thermodynamic cost of cruciform formation. This will reduce the kinetic barrier to cruciform extrusion by stabilising the protocruciform transition state. Extrusion is therefore accelerated by a lower free energy of activation of the transition state ($\Delta G^{\ddagger}$); this is illustrated in Fig. 8.15. While acceleration of cruciform extrusion by two-base loop formation may be responsible for the extreme undermethylation of DRL179, this has yet to be confirmed by *in vitro* studies of cruciform extrusion kinetics.

Figure 8.15

Acceleration of cruciform extrusion by protocruciform transition state stabilisation.



In summary, this chapter shows that central sequence composition of a palindrome profoundly affects the efficiency of Dam methylation at its centre *in vivo*. The high levels of methylation inhibition observed in both N2364 and JC7623 hosts are best explained by the formation of a persistent cruciform structure. While three of the four palindromes show an inverse correlation between the T_{MN} of the central 10 bp and the degree of undermethylation as predicted by *in vitro* studies, the sequence in DRL179 behaves anomalously. This may be due to the formation of two-residue cruciform loops by this palindrome centre.

CHAPTER 9

EFFECT OF PALINDROME CENTRE ASYMMETRY ON DAM METHYLATION *IN VIVO*

Introduction

Central interruptions to the inverted symmetry of a palindrome can overcome the inviability associated with such sequences (Collins, 1982; Mizuuchi *et al.*, 1982b). An asymmetric insertion of 57 bp has been determined to be the minimum required for an alleviation of lethality (Warren & Green, 1985). However, smaller lengths of central asymmetry may have intermediate effects: Chalker *et al.* (1993) examined the consequences of asymmetric insertions of 0 to 27 bp on the plating of λ phage bearing a long perfect palindrome. It was found that the severity of inviability declines with an increasing length of asymmetry, and that this hierarchy in plating is independent of the host genotype. These results implicate a centre-dependent reaction such as cruciform extrusion in the phenotypic effects of palindromic sequences. The effect of a 10 bp asymmetric insertion on undermethylation of the palindrome centre *in vivo* was therefore examined.

Results

Insertion of 10 bp Asymmetry at Palindrome Centre

Two non-palindromic oligonucleotides were cloned in the central *Sac*I site of the long DNA palindrome in DRL167 to create a 10 bp interruption.

Asymmetric Oligonucleotide with Central GATC Site

The oligonucleotides are complementary and can therefore form the double stranded species shown below. They share a number of features with the sequences described in Chapters 4 and 8: they are 14 nucleotides long, have 3' terminal *Sac*I linker sequences, and contain a central GATC target site. This GATC sequence additionally forms part of an *Xho*II target site (pur·GATC·pyr).

```
                Dam
5'      GTGGATCTCAAGCT 3'
3' TCGACACCTAGAGT      5'
   SacI     XhoII     SacI
```

However, unlike the insertions previously described, the oligonucleotides shown above are not self-complementary. Although the central GATC target is palindromic, this sequence would not be expected to form a stem-loop structure on its own. The insertion of these oligonucleotides therefore constitutes an asymmetric

interruption of 10 bp at the centre of the palindrome. The oligonucleotides were designed with a normal base composition in mind: the T_{MN} of the central 10 bp of the insert was calculated to be 74.2°C, which is very close to the T_{MN} of DNA of an average (50% G/C) base composition (73.6°C).

Cloning of Oligonucleotides in Palindrome Centre

The two oligonucleotides were combined in equimolar amounts, heated to 100°C and allowed to cool slowly to favour the formation of the double-stranded species. They were then cloned in the central *Sac*I site of λ DRL167 as described previously (Chapters 4 and 8).

Selection of λ Phage with Oligonucleotide Insertion in Palindrome

The *in vitro* packaged phage were amplified by one round of plating on JC7623 (*recBC sbcBC*). 45 plaques were picked into phage buffer and spotted onto bacterial lawns of N2361 (*rec*$^+$) and JC7623. Six phage growing on JC7623 but not on N2361 were selected and purified by two further rounds of growth on JC7623. Single plaques were picked for the preparation of phage stocks by plate lysates, again using JC7623 as a host strain.

λ DNA was prepared from the plate lysates, digested to completion with *Eco*RI and radiolabelled with [α^{35}S]dATP and Klenow enzyme. After phenol-chloroform and chloroform extraction and ethanol precipitation, the DNA was resuspended in the appropriate buffer and digested to completion with *Xho*II. The DNA was then extracted with phenol-chloroform and chloroform and precipitated with ethanol again, resuspended in formamide-EDTA gel loading buffer and electrophoresed on a 6% denaturing polyacrylamide gel. (Agarose gel electrophoresis of an *Xho*II restriction digest would have been impractical, as there are 21 *Xho*II sites in wild-type λ.) Five of the six phage contained a novel *Xho*II site at the palindrome centre; one was chosen and designated DRL180.

Analysis of *in Vivo* Dam Methylation of DRL180

In Vivo Methylation by Growth of DRL180 in N2364 and JC7623 Hosts

DRL180 was used to infect N2364 (*sbcC dam*$^+$) and JC7623 (*recBC sbcBC dam*$^+$) at a MOI of 0.1, and grown by liquid lysis. After complete cell lysis, phage particles were purified and λ DNA extracted using the large scale method.

Analysis of Methylation at GATC Sites in DRL180

The extent of Dam methylation at the palindrome centre was assayed by *Dpn*I and *Mbo*I cleavage as described in Chapter 4; [α^{35}S]dATP was used to radiolabel the *Eco*RI termini as described in Chapter 5. *Xho*II was used in the place of *Bam*HI to cleave the palindrome centre to completion. The samples were electrophoresed as described in Chapter 4 and autoradiographed to provide the image in Fig. 9.1 or exposed to a storage phosphor screen for the quantification of band intensities.

Quantification of Methylation at GATC Sites

The bands in lanes corresponding to *Dpn*I and *Mbo*I cleavage in Fig. 9.1 were quantified by PhosphorImager™ and their values are shown in Table 9.1. The level of Dam methylation at GATC sites *in vivo* and the relative efficiency of *Dpn*I and *Mbo*I cleavage at these sites were calculated as for Table 4.1, and are also shown in Table 9.1. The latter values are plotted in Figs. 9.2 and 9.3. The analysis of Dam methylation was repeated, yielding very similar results to those shown here.

Figure 9.1

Autoradiograph of Restriction Digest at GATC Sites in DRL180

Autoradiograph showing restriction digests of λ DRL180 DNA determining the level of Dam methylation at various GATC sites *in vivo*. Two sets of digests are shown, representing methylation in different *E. coli* hosts. The λ phage had been grown in either N2364 (*sbcC*) or JC7623 (*recBC sbcBC*) and DNA purified as described in the text. The DNA was cleaved with *Eco*RI and radiolabelled with [α^{35}S]dATP as described in the text, before cleavage with *Xho*II, *Dpn*I or *Mbo*I. The *E. coli* host strain and the appropriate enzymes are indicated above each lane. The full-length and half-length palindrome, and the control fragments (see Fig. 4.3) are shown by the labels on the right of the autoradiograph.

From left to right, the lanes are:

1. λ DRL180 [N2364 (*sbcC*) host] *Eco*RI *Xho*II
2. λ DRL180 [N2364 (*sbcC*) host] *Eco*RI *Dpn*I
3. λ DRL180 [N2364 (*sbcC*) host] *Eco*RI *Mbo*I
4. λ DRL180 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Xho*II
5. λ DRL180 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Dpn*I
6. λ DRL180 [JC7623 (*recBC sbcBC*) host] *Eco*RI *Mbo*I

Cleavage by *Xho*II at the central pur·GATC·pyr site generates the half-length palindrome. Digestion with *Dpn*I indicates the level of fully methylated DNA, and *Mbo*I the level of fully unmethylated DNA.

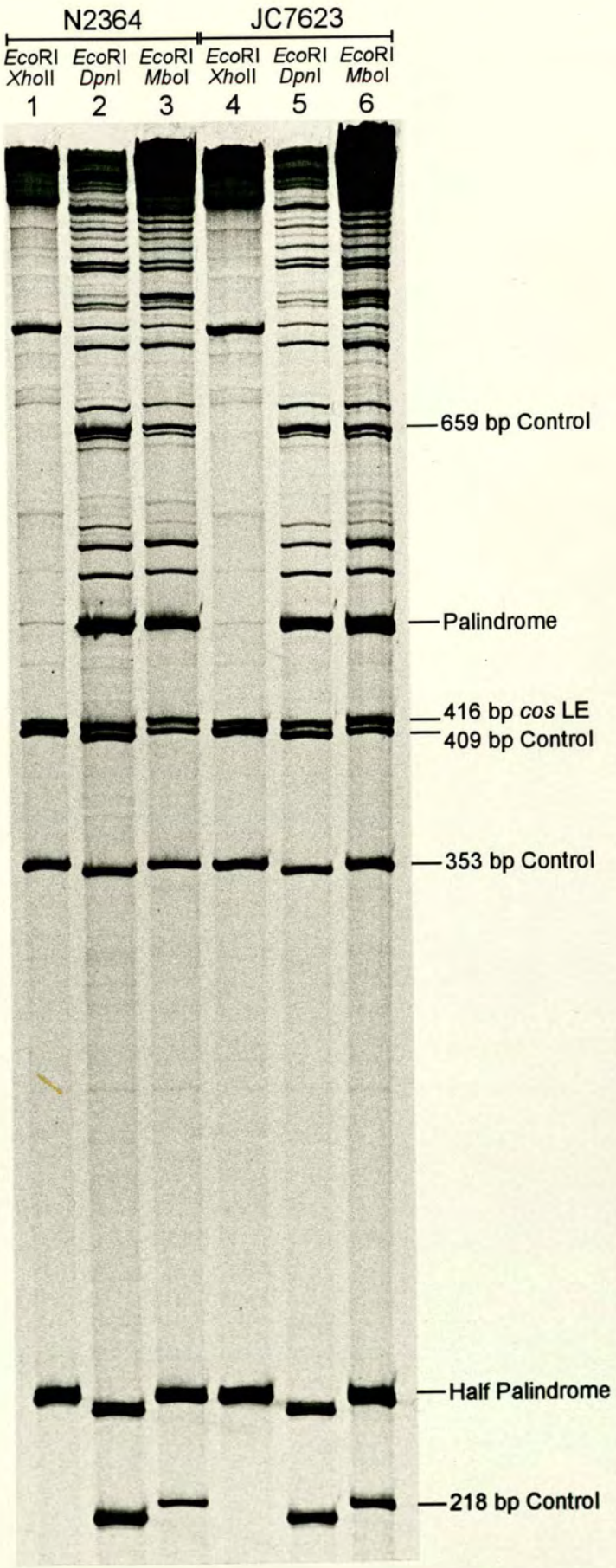

N2364
JC7623
EcoRI XhoII
EcoRI DpnI
EcoRI MboI
EcoRI XhoII
EcoRI DpnI
EcoRI MboI
1
2
3
4
5
6
659 bp Control
Palindrome
416 bp cos LE
409 bp Control
353 bp Control
Half Palindrome
218 bp Control

Table 9.1

Methylation in N2364 (*sbcC*) of GATC Sites in DRL180

The bands shown in lanes 2 and 3 of Fig. 9.1, which correspond to methylation in N2364 (*sbcC*), were quantified using a PhosphorImager™. Their intensities are displayed in Table 9.1 in the rows designated "Data". The quantified values (Uncorrected Data) for the full-length and half-length palindrome were halved (Corrected Data). The % cleavage at the palindrome centre by *Dpn*I or *Mbo*I was calculated using equation (9) and the % methylation at the palindrome centre was determined using the level of methylated DNA as measured by both *Dpn*I and *Mbo*I cleavage; the values from the two enzymes were averaged using equation (10). The average % cleavage at control sites was determined by equation (11) using the sum of the palindrome bands as a standard, and the average % methylation of the control sites was similarly calculated using equation (12). The bottom portion of the table shows the relative efficiency of cleavage by *Dpn*I and *Mbo*I at the palindrome and control GATC sites. For a detailed explanation of these calculations see Table 4.1

Table 9.2

Methylation in JC7623 (*recBC sbcBC*) of GATC Sites in DRL180

The bands shown in lanes 5 and 6 of Fig. 9.1, which correspond to methylation in JC7623 (*recBC sbcBC*), were quantified using a PhosphorImager™. Their intensities are displayed in Table 9.2 in the rows designated "Data". The values derived from these data were calculated as for Table 9.1 (see above).

Table 9.1

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$		*Mbo*I (GATC)
Uncorrected Data	Palindrome	76633		55918
	½ Palindrome	26884		44492
Corrected Data	Palindrome	38317		27959
	½ Palindrome	13442		22246
% Cleavage at Palindrome Centre		26.0		44.3
% Methylation of Palindrome Centre			40.8	
Control Data	659 bp Control	28709		6678
	409 bp Control	27086		7836
	353 bp Control	24287		17652
	218 bp Control	36343		6329
	Average Control	29106		9624
Average % Cleavage at Control Sites		56.2		19.2
Average % Methylation of Control Sites			68.5	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	46.2		231.2
	669 bp Control	98.6		69.4
	409 bp Control	93.1		81.4
	353 bp Control	83.4		183.4
	218 bp Control	124.9		65.8
	Average Control	100		100

Table 9.2

Restriction Endonuclease		*Dpn*I ($G^{Me}ATC$		*Mbo*I (GATC)
Uncorrected Data	Palindrome	58371		89547
	½ Palindrome	18007		97485
Corrected Data	Palindrome	29186		44774
	½ Palindrome	9004		48743
% Cleavage at Palindrome Centre		23.6		52.1
% Methylation of Palindrome Centre			35.7	
Control Data	659 bp Control	19982		12854
	409 bp Control	21306		16082
	353 bp Control	17772		39392
	218 bp Control	25582		16968
	Average Control	21161		21324
Average % Cleavage at Control Sites		55.4		22.8
Average % Methylation of Control Sites			66.3	
Relative Cleavage at GATC Sites (% of Average)	½ Palindrome	42.6		228.6
	669 bp Control	94.4		60.3
	409 bp Control	100.7		75.4
	353 bp Control	84.0		184.7
	218 bp Control	120.9		79.6
	Average Control	100		100

Figure 9.2

Relative Cleavage by *Dpn*I at GATC Sites in DRL180

Graph showing the relative efficiency of cleavage by *Dpn*I at GATC sites in λ DRL180 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨). The values plotted are those shown on the left-hand side of the bottom portions of Tables 9.1 and 9.2. The intensities of the bands shown in Fig. 9.1 are presented relative to the average intensity of the control bands, although the value for the half-length palindrome band was halved to correct for radiolabelling at two *Eco*RI sites. Error bars are also shown, representing the range of relative cleavage by *Dpn*I at GATC sites as calculated from individual restriction digests. Figure 9.2 represents graphically the presence of Dam methylation at various GATC sites in DRL180.

Figure 9.3

Relative Cleavage by *Mbo*I at GATC Sites in DRL180

Graph showing the relative efficiency of cleavage by *Mbo*I at GATC sites in λ DRL180 following Dam methylation *in vivo* in N2364 (*sbcC*) (▩) or JC7623 (*recBC sbcBC*) (▨). The values plotted are those shown on the right-hand side of the bottom portions of Tables 9.1 and 9.2. The data are presented as in Fig. 9.2. Error bars are also shown, representing the range of relative cleavage by *Mbo*I at GATC sites as calculated from individual restriction digests. Figure 9.3 represents graphically the absence of Dam methylation at various GATC sites in DRL180.

Figure 9.2

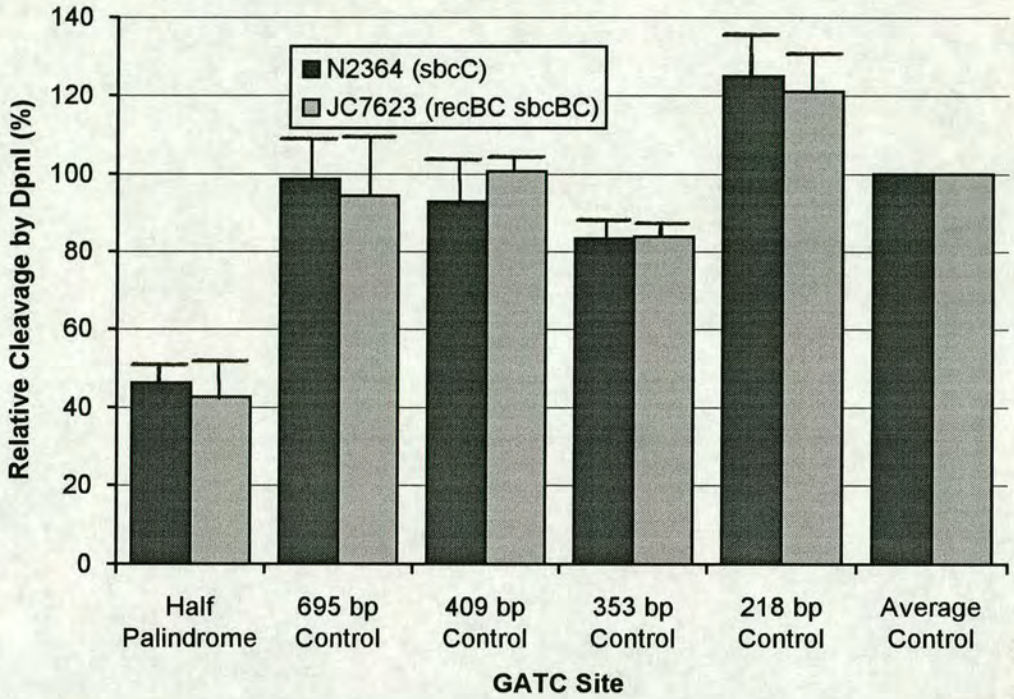
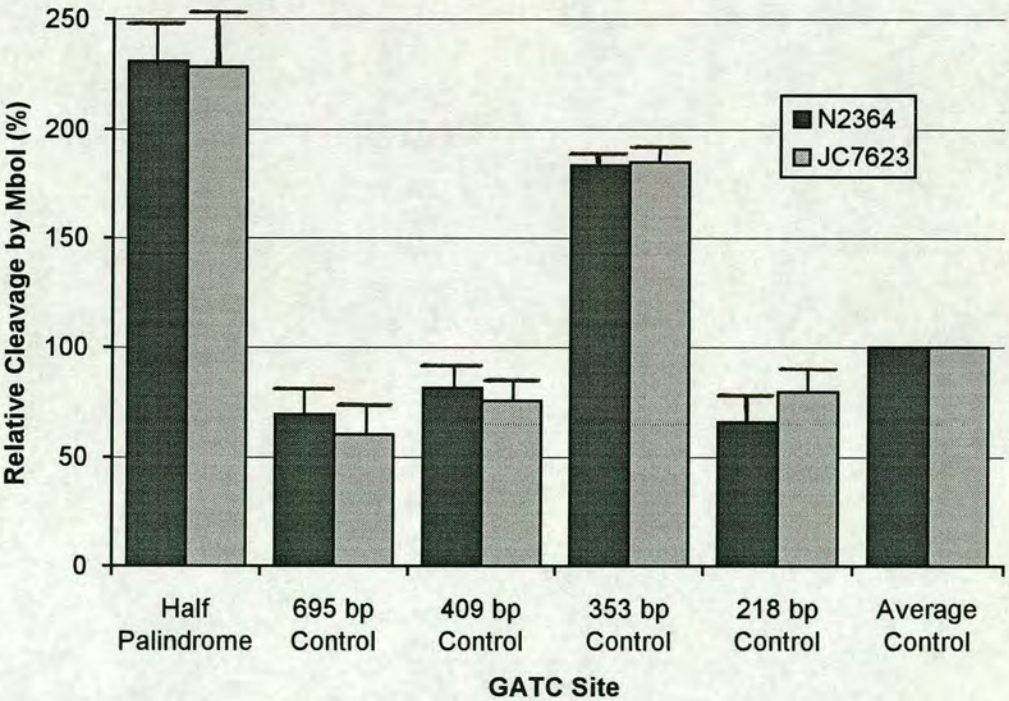


Figure 9.3



Conclusion

The results presented in this chapter demonstrate that the 10 bp asymmetric palindrome centre in DRL180 shows unexpectedly high levels of methylation inhibition *in vivo*. Analysis of this methylation inhibition by kinetic modelling (as set out in Chapter 8) indicates that the palindrome centre of DRL180 is inaccessible to Dam methylase for 55% of the λ lytic lifecycle in N2364, and for 59% of the time in JC7623. If *in vivo* cruciform extrusion is responsible for this methylation inhibition, then it follows that the asymmetric palindrome in DRL180 extrudes more often than the perfect palindrome in DRL176 (see Fig. 8.13).

This is clearly unlikely; calculations (Hsieh & Wang, 1975; Warren & Green, 1985) indicate that an asymmetry of 20 bp would impart upon a palindromic sequence an insurmountable kinetic barrier to cruciform extrusion under physiological conditions. Hence there are four possible explanations for the results described here:

1. cruciform extrusion does not occur *in vivo*, and all of the instances of undermethylation reported so far are due to alternative mechanisms of methylation inhibition;
2. cruciform extrusion by perfect palindromes occurs *in vivo*, and alternative mechanisms of methylation inhibition act specifically on DRL180;
3. cruciform extrusion by perfect palindromes occurs *in vivo*, and alternative mechanisms of methylation inhibition act on both the perfect and asymmetric palindromes;
4. cruciform formation occurs *in vivo*, but is underestimated by the methylation assay in the case of the perfect palindromes.

Although the first possibility has been discussed in the previous chapter, it is prudent to consider here whether an effect of sequence context on Dam methylase activity is responsible for undermethylation of the palindrome centres. In particular, an extreme instance of this effect at the palindrome centre of DRL180 would be compatible with the second possibility described above. Analysis of *in vitro* Dam methylation (as described in Chapter 6) using λ DNA from DRL176 – 180 would dispel any such uncertainty.

The second possibility invokes a mechanism that specifically recognises the asymmetric palindrome centre of DRL180. This mechanism may involve protein binding at a serendipitously-formed recognition site, and could easily be tested using

an asymmetric insertion with a different sequence. An intriguing alternative is the stimulation of genetic recombination by single-strand loops. This idea has previously been presented as part of a molecular mechanism for genetic recombination (Sobell, 1972). Although such models are currently not favoured, they might account for the high levels of methylation inhibition observed in DRL180. This palindrome may form hairpins with single-stranded loops of 10 nucleotides, and consequently stimulate recombination more than the four (or two) base cruciform loops formed by the perfect palindromes. This recombination may in turn prevent reabsorption of the hairpin and methylation of the palindrome centre. However, these ideas are very speculative and require further investigation, and will therefore not be discussed any further.

The third and fourth possibilities are probably the best explanations for the results presented in this chapter. However, they are considered in the Discussion at the end of this thesis and will therefore not be examined here.

The effects of central asymmetry on the propagation of palindromic DNA in λ have been shown to be consistent with cruciform extrusion *in vivo* (Chalker *et al.*, 1993). It would be predicted that the plating behaviour of DRL180 reflect the extent of cruciform formation by its resident palindrome. This independent measure of the biological effects of palindromic DNA would therefore assist in evaluating the four possibilities listed above. Such an examination of the viability of DRL176 – 180 is presented in Chapter 10.

In summary, this chapter demonstrates that the asymmetric palindrome centre in DRL180 shows a high level of methylation inhibition *in vivo*. This is inconsistent with studies of cruciform formation by palindromes with central asymmetric interruptions, and therefore compromises earlier conclusions about cruciform extrusion *in vivo*.

CHAPTER 10

EFFECT OF PALINDROME CENTRAL SEQUENCE ON λ VIABILITY

Introduction

Mutation of the *sbcC* gene is necessary and sufficient to permit the stable propagation of DNA replicons containing long DNA palindromes (Chalker *et al.*, 1988). Nevertheless, λ derivatives carrying such palindromic sequences exhibit a reduced viability in *sbcC* hosts when compared to phage without a palindrome. The presence of asymmetric inserts at the palindrome centre increases the viability of these phage, suggesting that a centre-dependent pathway for the phenotypic effects of palindromic DNA exists in *sbcC* as well as wild-type *E. coli* (Chalker *et al.*, 1993). A recent study of the plating behaviour of λ phage on *sbcC* mutants showed clear differences in viability that were attributable to changes in the central sequence of their long perfect palindromes (Davison & Leach, in the press). A similar analysis of the effect of palindrome central sequence on the viability of λ DRL176 – 180 in *E. coli sbcC* hosts was therefore undertaken.

Results

Plating Behaviour of DRL176 – 180 on N2364

The phenotype conferred upon DRL176 – 180 by the central sequence changes to the long DNA palindrome were assessed by plating on a lawn of N2364 (*sbcC*), where differences in plaque size between the various λ phage are most visible.

Optimisation of Plating Conditions

The plating conditions were optimised in order to accentuate differences in the viability of DRL176 – 180. The plaque sizes were initially assessed on BBL plates with a standard NaCl concentration of 5g l^{-1}. The NaCl concentration was then varied, as higher concentrations have been found to amplify plaque size differences (Davison & Leach, in the press). The NaCl concentration was eventually adjusted to 14g l^{-1}; this level of salt was found to augment the differences in viability most.

While this effect of increased NaCl concentration may be due to a non-specific reduction in λ viability (which is most apparent in phage with poor viability), it is possible that the viability differences are the consequence of elevated levels of negative supercoiling. Environmental stimuli such as osmotic stress lead to variations in DNA supercoiling, which in turn regulate gene expression in *E. coli* (Higgins *et al.*, 1988; reviewed by Higgins *et al.*, 1990a). The increased negative supercoiling in *E.*

coli host cells grown on media with a high NaCl concentration may result in more cruciform extrusion. If cruciform formation is responsible for the reduced viability of palindrome-bearing phage, then higher levels of extrusion would account for the amplified viability differences seen on media with 14 g l^{-1} NaCl.

The use of Select® trypticase (Becton Dickinson) was found to benefit plating non-specifically and was consequently used. Further details of the optimised plating conditions are described in the Materials & Methods (Chapter 2).

Quantification of Plating Behaviour of DRL176 – 180

Single plaques of DRL176 – 180 and DRL152 were picked into phage buffer. DRL152 (*spi6 cI*857 $\chi^{+}C153$) was chosen as a control, being isogenic to DRL176 – 180 apart from lacking the palindrome and the ΔB deletion. Around 200 – 300 PFU of each phage were incubated with N2364 at 37°C for 20 minutes, and then plated on the media described above. These conditions were chosen to maximise preadsorption of the phage. After overnight incubation at 37°C, the area of individual plaques on the cell lawn was measured accurately using a Quantimet 970 Image Analyser. Approximately 50 – 200 plaques per plate were analysed, using 9 plates for each λ strain. Since plaque area is an indirect product of the λ burst size, this form of analysis provides a measure of the phage viability. The quantified plaque areas are shown in Figs. 10.1 and 10.2. The first graph (Fig. 10.1) shows frequency distributions of plaque areas produced by the four phage with perfect palindromes, while the second graph (Fig. 10.2) shows cumulative frequency distributions of plaque areas produced by all six phage. The latter form of presentation was chosen for clarity.

The mean plaque size produced by a λ strain is not an ideal description of the distributions shown in Figs. 10.1 and 10.2, as it is susceptible to outliers. Outliers are relatively common in this analysis for two reasons:

1. λ phage that are initially unabsorbed produce pinprick plaques on the cell lawn,
2. and phage that have lost the palindrome produce very large plaques.

A 10% trimmed mean was therefore calculated from the data for each phage. This is shown in Table 10.1. However, the frequency distributions also exhibit considerable skewness, to which the mean is sensitive. The median and interquartile range are robust statistics and were therefore chosen to describe the plaque area distributions; their values are also presented in Table 10.1.

Figure 10.1

Frequency Distributions of DRL176 – 179 Plaque Sizes on N2364 Lawn

Frequency distributions of plaque areas produced by DRL176 – 179 plating on an N2364 cell lawn. The plaque areas were quantified using image analysis as described in the text. The frequency distributions of DRL180 and DRL152 have been omitted for clarity. DRL176 – 179 carry perfect palindromes with differing central sequences.

Figure 10.2

Cumulative Frequency Distribution of DRL176 – 180 and DRL152 Plaque Sizes on N2364 Lawn

Cumulative frequency distributions of plaque areas produced by DRL176 – 180 and DRL152 plating on an N2364 cell lawn. The plaque areas were quantified using image analysis as described in the text. DRL180 carries a palindrome with an asymmetric 10 bp central insertion. DRL152 does not carry a palindrome.

Figure 10.1

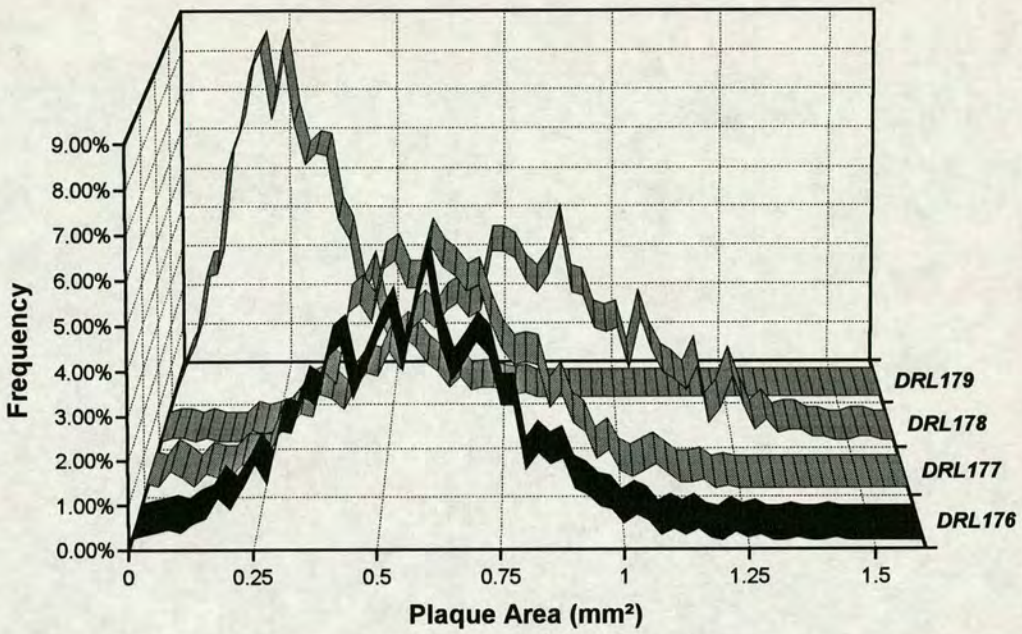


Figure 10.2

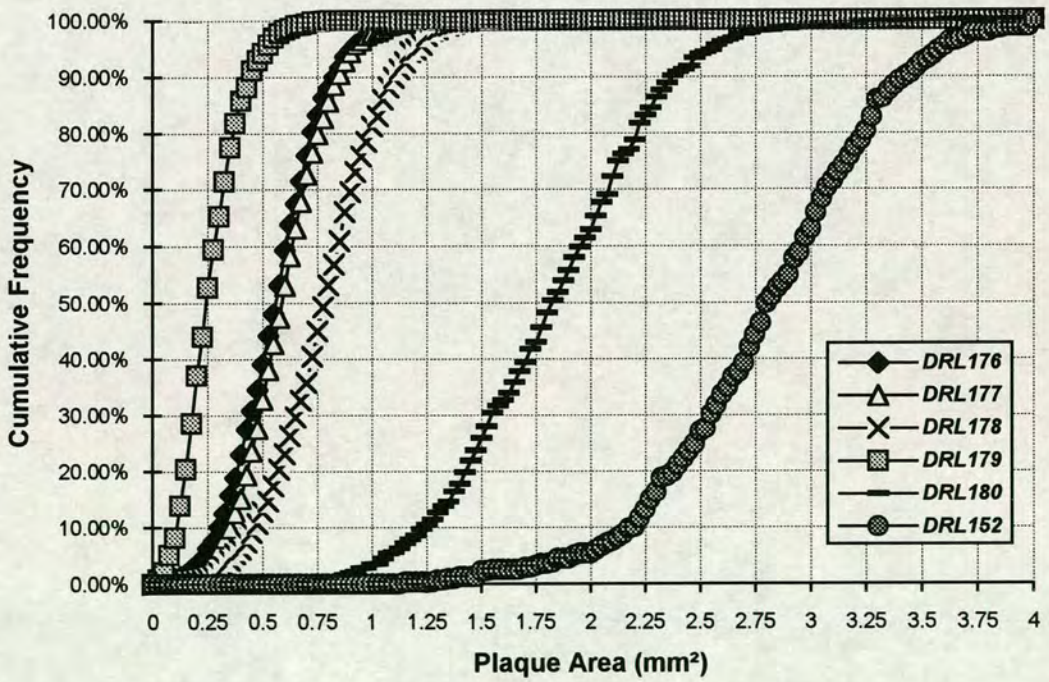


Table 10.1

Viability of DRL176 – 180 and DRL152 as Measured by Plaque Area

Table showing the 10% trimmed mean, median and interquartile range of the plaque areas produced by plating of DRL176 – 180 and DRL152 on an N2364 cell lawn. A trimmed mean was calculated to minimise the effect of outliers. The median differs from the 10% trimmed mean in some cases due to the skewness of the plaque area distributions; the median is resistant to skewness.

Table 10.1

λ Strain	Palindrome Central Sequence	Plaque Area (mm^2)		
		10% Trimmed Mean	Median	Interquartile Range
DRL176	5' GT**G**GATC**C**AC 3' 3' CA**C**CTAG**G**TG 5'	0.581	0.585	0.437 – 0.719
DRL177	5' GT**T**GATC**A**AC 3' 3' CA**A**CTAG**T**TG 5'	0.614	0.614	0.483 – 0.740
DRL178	5' GT**A**GATC**T**AC 3' 3' CA**T**CTAG**A**TG 5'	0.807	0.804	0.642 – 0.965
DRL179	5' GT**C**GATC**G**AC 3' 3' CA**G**CTAG**C**TG 5'	0.280	0.268	0.190 – 0.364
DRL180	5' **GTG**GATC**TCA** 3' 3' **CAC**CTAG**AGT** 5'	1.845	1.848	1.517 – 2.149
DRL152	N/A	2.837	2.834	2.502 – 3.178

Effect of *ruvC* and *recG* Mutations on λ Viability

The *E. coli ruvC* gene encodes a Holliday junction resolvase which cleaves four-way DNA junctions (Connolly *et al.*, 1991; Iwasaki *et al.*, 1991), and the product of the *recG* gene has been shown to specifically dissociate synthetic Holliday junctions *in vitro* (Lloyd & Sharples, 1993a). Consequently, both of these proteins are capable of interacting with extruded cruciforms *in vivo*. The effect of *ruvC* and *recG* mutations on the viability of DRL176 – 180 and DRL152 phage was therefore examined by the plaque area assay described above.

λ DRL176 – 180 and DRL152 were plated on lawns of LR299 (*sbcC ruvC*) and LR327 (*sbcC ruvC recG*) (L. Ryder & R. Lloyd, unpublished) as described earlier in this chapter. Although an increase in plaque size was observed, this effect was non-specific; phage with perfect, imperfect and no palindrome were affected similarly. It was therefore concluded that the RuvC and RecG proteins are not responsible for the reduced viability of DRL176 – 179 in *sbcC* hosts. However, this does not imply that the palindromic sequences in these phage are incapable of undergoing cruciform extrusion *in vivo*, as it is uncertain whether RuvC or RecG would cleave a cruciform structure. Furthermore, the product of the *rus* gene has recently been identified as a Holliday junction resolvase (R Lloyd, unpublished observations); this redundancy in resolvase activities undermines a simplistic genetic approach such as the one presented here.

Conclusion

The results of the plaque area assay presented here show clearly that long DNA palindromes compromise the viability of λ phage in an *sbcC* host. This reduction in viability is dependent upon the presence of asymmetry at the palindrome centre: DRL180, which has a 10 bp interruption to the palindromic symmetry, produces considerably larger plaques on an N2364 lawn than DRL176 – 179, which carry perfect palindromes. If the reduced viability of the latter four phage is primarily due to cruciform extrusion, then it follows that the inverted repeat in DRL180 does not form a cruciform structure very often. This conclusion, which confirms the results of Chalker *et al.* (1993), stands in stark contrast to the unexpected results described in the previous chapter.

The plaque area assay also shows that base sequence changes at the centre of a perfect palindrome can affect the viability of its carrier replicon; DRL176 – 179 produce plaques of varying sizes on an N2364 lawn. While this result may appear to

be in agreement with the findings presented in Chapter 8, a closer inspection of the data reveals some inconsistencies. The hierarchies of undermethylation and plaque size observed for these four phage are shown below:

Undermethylation: DRL179 > DRL178 > DRL177 > DRL176

Plaque Size: DRL179 < DRL176 ≈ DRL177 < DRL178

Note that decreased plaque size is analogous to increased undermethylation. In this respect, only DRL179 behaves consistently in both assays. However, DRL179 has the highest predicted T_{MN} of the central 10 bp and would therefore be expected to undergo cruciform extrusion *in vivo* least frequently. This forces the conclusion that either one or both of the assays presented here is not an exclusive measure of cruciform extrusion *in vivo*.

In Chapter 8 it was argued that undermethylation of the palindrome centres of DRL176 – 178 correlates with the T_{MN} of the central 10 bp, and that the palindrome centre of DRL179 shows the greatest undermethylation due to possible formation of a two-base cruciform loop *in vivo*. Conversely, the plaque area data shows no correlation whatsoever with the T_{MN} of the central 10 bp, even if DRL179 is ignored. The correlation of plaque area with the stability of the protocruciform transition state as determined by the FOLD program (Table 8.11) is also poor. However, FOLD predicts that the loops formed by DRL177 and DRL178 contain six nucleotides. If all loops (including DRL179) are assumed to consist of four bases and their energy of formation is ignored, then differences in hairpin stability will depend on the composition of the base pairs in the stem. The doublet immediately preceding the loop constitutes the only difference between the hairpin stems, and is shown in bold with its predicted T_{MN} (Gotoh & Tagashira, 1981) in Table 10.2.

Table 8.11

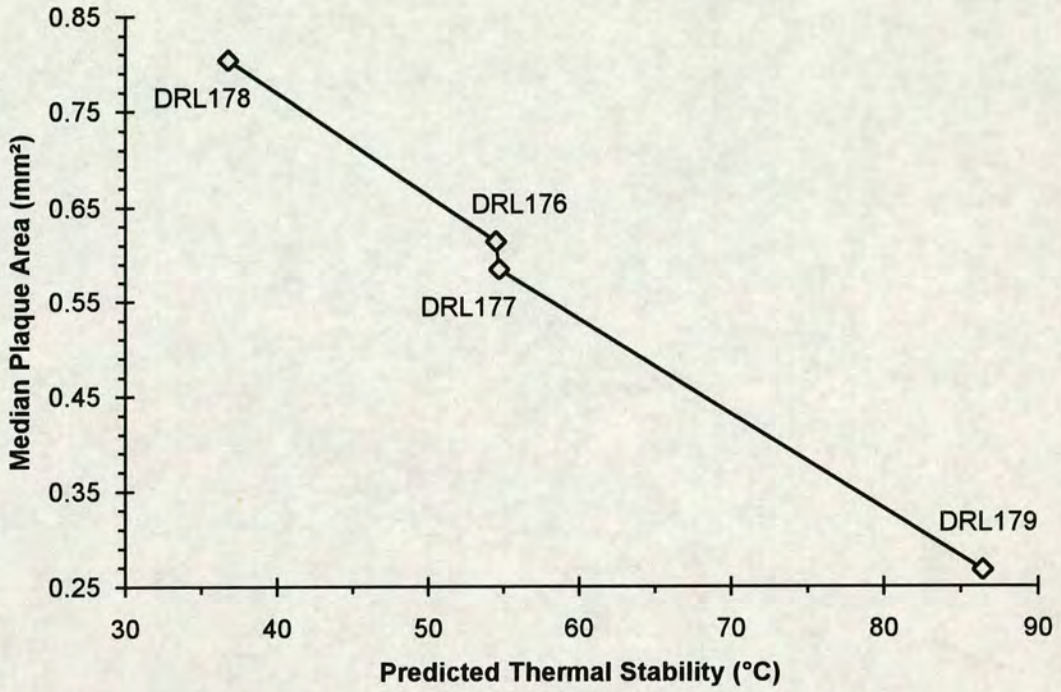
Hairpins with four-base loops and predicted T_{MN} values of the differing doublet.

λ Strain	DRL176	DRL177	DRL178	DRL179
Hairpin	G 5'~TG**TG** A 3'~AC**AC** T C	G 5'~TG**TT** A 3'~AC**AA** T C	G 5'~TG**TA** A 3'~AC**AT** T C	G 5'~TG**TC** A 3'~AC**AG** T C
T_{MN} of Doublet	54.71	54.50	36.73	86.44

These values are also shown in Fig. 10.3, which shows the relationship between the median plaque area and the predicted T_{MN} of the differing doublet.

Figure 10.3

Median plaque area produced on an N2364 lawn vs. predicted thermal stability (T_{MN}) of the terminal base pair doublet in the hairpin stem.



This relationship suggests that the stability of the hairpin stem determines the viability of the carrier phage. While such a correlation may have arisen fortuitously, it does not lead to unrealistic conclusions. Fig. 1.4 (see Introduction to thesis) presents a scheme for palindrome-mediated replicon inviability wherein hairpin structures present on the lagging strand of replication prevent efficient DNA synthesis, and are consequently degraded by the SbcCD protein. These hairpins would also impede DNA replication in an *sbcC* mutant, accounting for the reduced viability of palindrome-bearing phage in an *sbcC* host. If the ability of DNA polymerase to replicate such hairpins depends in part upon the thermal stability of the stem, then base pair changes such as those in DRL176 – 179 would affect the efficiency of DNA replication and therefore the viability of the vector.

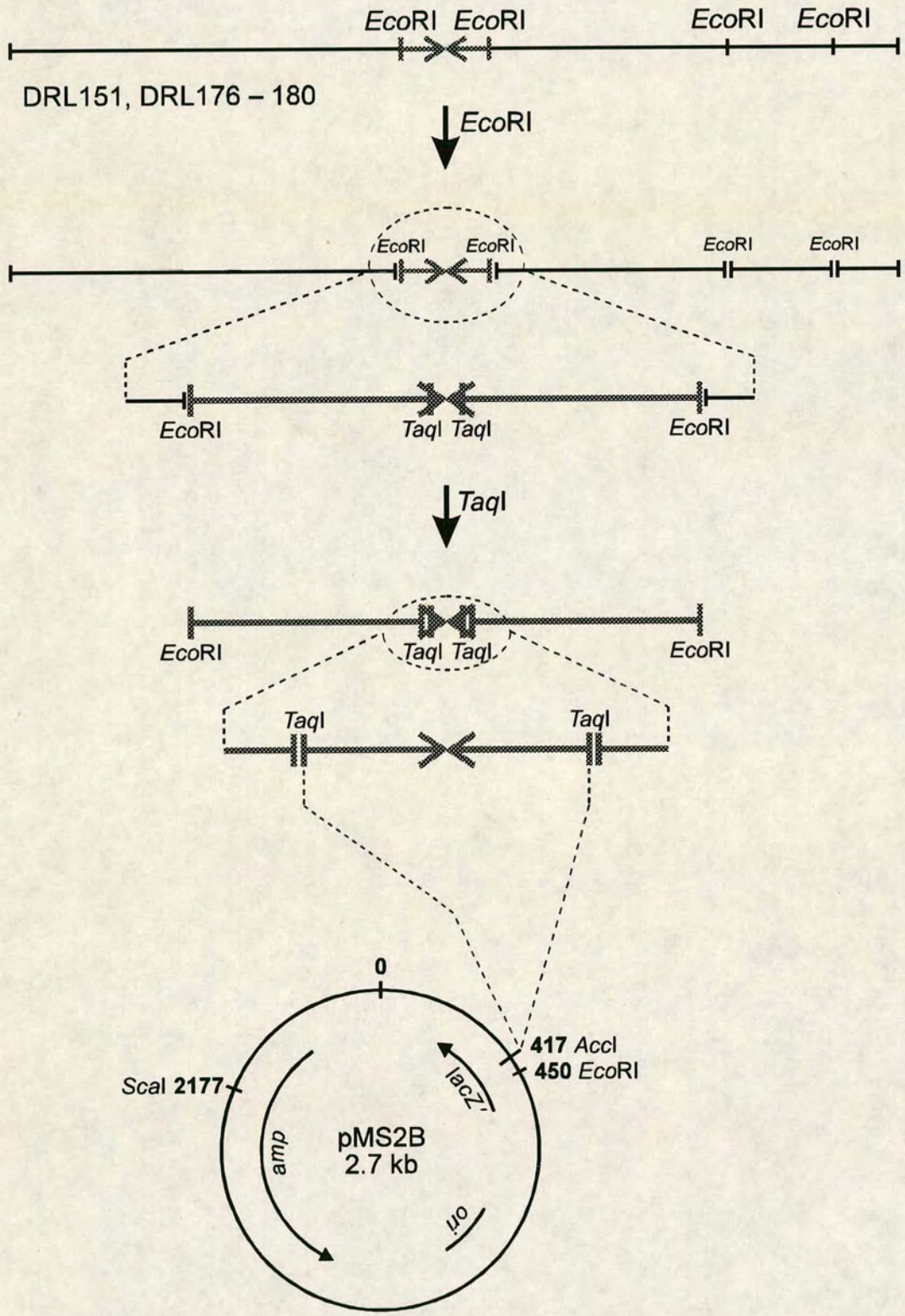
In summary, this chapter shows that changes to the central sequence of a long DNA palindrome affect the viability of its λ vector, and that the presence of asymmetry at the palindrome centre greatly increases this viability. The viability of DRL176 – 179 correlates with the T_{MN} of the terminal base pair doublet in the hairpin stem.

CHAPTER 11

SUBCLONING AND DNA SEQUENCING OF PALINDROME CENTRES

Figure 11.1

Subcloning of Palindrome Centres in pMS2B (see text for details).



Introduction

The conclusions of the preceding chapters are critically dependent upon the veracity of the DNA sequence changes introduced at the palindrome centres. Although the oligonucleotide inserts at the palindrome centre can be detected by restriction digest with the appropriate endonuclease, it remains necessary to verify the actual DNA sequence of these insertions.

However, sequencing of full-length palindromes has proved problematic; the centre of the palindrome appears to be particularly refractory to a wide variety of DNA sequencing techniques (A. Davison, unpublished observations). This is presumably due to formation of hairpins at the primer annealing step. DNA sequencing at a high temperature using *Taq* DNA polymerase does not overcome this problem (A. Davison, unpublished observations).

The centre of the palindrome was therefore subcloned in pMS2B, a derivative of plasmid pUC18 (Norrander *et al.*, 1983; Yanisch-Perron *et al.*, 1985), which for unknown reasons stabilises long palindromes (C. Blake, M. Shaw & D. Leach, unpublished observations). This 28 bp fragment containing the palindrome centre does not pose any serious problems to DNA sequencing. The subcloning of the palindrome centres in pMS2B is illustrated schematically in Fig. 11.1.

Results

Subcloning of Palindrome Centres

Around 5 µg of DRL151 and DRL176 – 180 DNA was digested to completion with *Eco*RI and electrophoresed on a 1.5% agarose gel. Bands of ~468 bp were excised from the gel and the palindromic DNA purified using the GENECLEAN® procedure. This DNA was digested to completion with *Taq*I, which cuts at TCGA giving a 28 bp fragment encompassing the palindrome centre (see Appendix). However, the palindrome centre of DRL179 contains the sequence:

```
5' ~AGCTGTCGATCGACAGCT~ 3'
3' ~TCGACAGCTAGCTGTCGA~ 5'
        TaqI  TaqI
```

Excision of the palindrome centre by *Taq*I cleavage was therefore impractical, as the restriction endonuclease would additionally cut at one of the tandem *Taq*I sites shown above. The palindrome centre of DRL179 was therefore excised using *Bst*BI, which

cuts at TTCGAA sequences. The *Taq*I sites flanking the 28 bp palindrome centre are also *Bst*BI sites (see Appendix), unlike the tandem *Taq*I sites at the palindrome centre of DRL179.

Plasmid pMS2B was digested to completion with *Acc*I, which cuts at GT(A/C)(G/T)AC; in pMS2B the target site is GTCGAC. Cleavage with *Acc*I produces a 5' CG overhang that is complementary to the DNA ends produced by *Taq*I digestion. The *Taq*I digest of the palindrome was purified using the MERmaid™ procedure and ligated with the cut pMS2B DNA. After ligation, the DNA was precipitated with ethanol, resuspended in the appropriate buffer and digested with *Acc*I. Insertion of the *Taq*I fragment inactivates the *Acc*I site; the second round of *Acc*I digestion therefore selects against pMS2B plasmids that have failed to clone the palindrome centre. The DNA was then used to transform DL733 (Δ*sbcCD*) to ampicillin resistance.

The *Acc*I site of pUC18 (and pMS2B) is located in the multiple cloning site, which is inserted into the 5' end of a portion of the *lacZ* gene (*lacZ'*). Insertion of the palindrome centre into the *Acc*I site will stop translation of the *lacZ* α-peptide from *lacZ'*, preventing complementation of the chromosomal *lacZ* deletion (Norrander *et al.*, 1983). As active β-galactosidase is required to form blue dye in the presence of Xgal (5-bromo-4-chloro-3-indolyl-β-D-galactoside), cells harbouring plasmids with an insertion of the palindrome centre into the *Acc*I site can be identified by white colonies on plates containing ampicillin and Xgal. Transformants were consequently spread and purified on such plates. Single colonies were cultured in L broth supplemented with ampicillin and plasmid DNA isolated using the small scale method.

Plasmid DNA was digested to completion with *Eco*RI and *Sca*I and run on a 1% agarose gel. These enzymes cut either side of the *Acc*I site to produce a fragment of 975 bp including the insertion (Fig. 11.1). Clones which appeared to have an insert of the correct size were selected; their appropriate *E. coli* hosts were cultured in L broth supplemented with ampicillin and plasmid DNA isolated using the QIAGEN method. The DNA was additionally purified using the GENECLEAN® procedure.

DNA Sequencing of Palindrome Centres

Around 3 μg of plasmid DNA was sequenced by the dideoxynucleotide chain-terminating method using the Sequenase® v2.0 kit as described in the Materials & Methods. The extension reactions were performed at 45°C to minimise secondary structure formation by the palindrome centre; band compression had been observed

after incubation at the normal temperature of 37°C. The samples were electrophoresed on a denaturing gradient polyacrylamide gel, which was subsequently dried and autoradiographed.

The sequences of all the plasmids with inserts of the palindrome centre were as predicted, and are shown below:

pDRL151 5' TCGAATGAGCTGT**G**GATC**C**ACAGCTCATTCGA 3'

pDRL176 5' TCGAATGAGCTGT**G**GATC**C**ACAGCTCATTCGA 3'

pDRL177 5' TCGAATGAGCTGT**T**GATC**A**ACAGCTCATTCGA 3'

pDRL178 5' TCGAATGAGCTGT**A**GATC**T**ACAGCTCATTCGA 3'

pDRL179 5' TCGAATGAGCTGT**C**GATC**G**ACAGCTCATTCGA 3'

pDRL180 5' TCGAATGAGCT**GTG**GATC**TCA**AGCTCATTCGA 3'

Conclusion

The sequences of the palindrome centres were all as expected. The conclusions of the previous chapters concerning the effects of central sequence changes are therefore valid.

The subcloned palindrome centres may be used in future work determining the rate of cruciform extrusion *in vitro*. This will hopefully shed light on the anomalous behaviour of DRL179: if the palindrome centre of this phage extrudes unusually rapidly *in vitro* then the case for cruciform extrusion *in vivo* by a centre-dependent pathway will be strengthened (see Chapter 8). The subcloned palindrome centres will also facilitate positive control experiments confirming the inhibition of Dam methylase activity by cruciform loops.

In summary, this chapter confirms that the palindrome central sequences are as predicted.

CHAPTER 12

DISCUSSION

SUMMARY

This thesis describes the development of a non-intrusive assay for cruciform DNA extrusion *in vivo*. Dam methylase was used to probe for the formation of a methylation-resistant structure at the centre of a long DNA palindrome in bacteriophage λ (Chapter 4). *Eco*RI methylase, which has been used successfully by others to detect Z-DNA *in vivo*, was not chosen due to possible incompatibilities with λ vectors (Chapter 3); such problems were not experienced with Dam methylase. The assay is internally controlled: target sites that are not expected to form a methylation-resistant structure are present on the same DNA molecule as the palindrome (Chapter 4). It is therefore possible to analyse the methylation inhibition by kinetic modelling (Chapter 8).

The centre of a long DNA palindrome in λ was found to be undermethylated *in vivo* (Chapter 4). This undermethylation was not significantly affected by the presence of a χ^+ site in the λ vector, or by growth of the phage in different (*recBC sbcBC* and *sbcC*) host strains (Chapter 5). Modification of relaxed λ DNA *in vitro* using purified Dam methylase shows that undermethylation of the palindrome centre is due specifically to the inhibition of Dam methylation *in vivo*, and is not the result of sequence context effects (Chapter 6).

The undermethylation of the palindrome centre is profoundly affected by its sequence composition (Chapter 8). This is in keeping with the results of *in vitro* experiments showing that the rate of cruciform extrusion is critically dependent on the palindrome central sequence. These observations have further established that the predicted thermal stability (T_{MN}) of the central ~10 bp of a palindrome correlates inversely with its rate of cruciform extrusion *in vitro* (Murchie & Lilley, 1987). A similar relationship between the predicted T_{MN} of the central 10 bp and the degree of undermethylation was found to exist for three of the four perfect palindromes studied (Chapter 8). However, the palindrome centre with the highest predicted T_{MN} (DRL179) exhibits the greatest degree of methylation inhibition *in vivo*, indicating that it undergoes cruciform extrusion most frequently. It is possible that this palindrome is capable of forming two-residue cruciform loops *in vivo*, which would account for its anomalous behaviour.

A palindrome was constructed with a central asymmetry of 10 bp (Chapter 9). This palindrome centre showed unexpectedly high levels of methylation inhibition *in vivo*, even exceeding those observed at a perfectly symmetrical centre. In contrast, the

λ phage carrying this palindrome shows a plating behaviour that is consistent with previous observations (Chalker *et al.*, 1993; Davison & Leach, in the press); the presence of a 10 bp asymmetry at the palindrome centre greatly increases the viability of the carrier phage (Chapter 10). Furthermore, the viability of λ vectors is not commensurate with the degree of methylation inhibition at perfect palindrome centres *in vivo*. However, a relationship between the plaque area produced by these phage and the predicted thermal stability of terminal base pairs in the hairpin stems can be seen (Chapter 10).

In summary, this thesis presents evidence consistent with extrusion of cruciform DNA *in vivo*. A centre-dependent pathway may contribute to cruciform formation, but evidence for this is complicated by the unexpected results obtained with an imperfect palindrome, and by the lack of a clear relationship between the effects of central sequence changes on methylation and λ viability.

DISCUSSION

The results presented in this thesis are not straightforward and may appear contradictory. This is unsurprising, given that the assays used are *in vivo* and indirect. The data is therefore open to alternative and conflicting interpretations. An effort has been made to discuss these ideas within the context of the relevant experiments, and they are presented in the Conclusion of each chapter. However, even after a great number of alternative hypotheses have been dismissed, the central deductions of this thesis remain at odds. This final Discussion is an attempt to unify these conclusions.

The results presented in this thesis suffer from three major inconsistencies:

1. the palindrome centre of DRL179 shows the highest level of methylation inhibition *in vivo*, and its carrier phage produces the smallest plaques on an *E. coli* lawn, in spite of having the highest predicted T_{MN} of its central 10 bp;
2. the palindrome centre of DRL180 shows substantial levels of methylation inhibition *in vivo* in spite of an asymmetric insertion of 10 bp;
3. with one exception, the levels of methylation inhibition observed at the palindrome centres *in vivo* is not commensurate with the (in)viability of their λ vectors.

The first of these problems has been discussed extensively in Chapter 8 and is incapable of resolution without further experiments such as the measurement of cruciform extrusion kinetics *in vitro* and an NMR study of hairpin loop structure.

The second and third inconsistencies listed above may be related in a manner that is suggested by the data in Chapter 10. The correlation between the plaque area produced by λ phage carrying perfect palindromes and the predicted thermal stability of the terminal base pair doublet in the hairpin stem (Fig. 10.3) indicates that S-type cruciform extrusion *per se* is not responsible for palindrome-mediated inviability. *In vitro* studies have not revealed any relationship between the rate of S-type cruciform extrusion and the stability of this base pair doublet. Conversely, the viability of the λ phage shows no relationship with the T_{MN} of the central 10 bp of the palindrome, which has been shown by *in vitro* studies to correlate inversely with the rate of S-type cruciform extrusion (Murchie & Lilley, 1987). Rather, the stability of the hairpin stem appears to determine the severity of the phenotype associated with palindromic DNA, namely its failure to replicate. The inability of DNA helicases to unwind stable stem-loop structures may lead to an arrest of replication fork progression, and thereby terminate DNA synthesis (Marians, 1992).

This hypothesis proposes that differences in the viability of phage with perfect palindromes are primarily the result of differences in hairpin stem stability. These structures may arise by cruciform extrusion, or by hairpin formation from single-stranded DNA. In the case of the asymmetric palindrome, the former pathway is kinetically forbidden. It is therefore tempting to propose that its carrier phage has an increased viability as a consequence of reduced levels of stem-loop structures. However, such a reduced incidence of stem-loop structures would be reflected in decreased methylation inhibition at the palindrome centre. Instead, comparable levels of methylation inhibition are seen at the centres of both perfect and asymmetric palindromes. This suggests that a DNA secondary structure is formed at the centre of the asymmetric palindrome, but that it does not compromise replication to the same degree as the DNA secondary structure formed at the centre of the perfect palindromes.

This disparity in the viability effects of stem-loop structures formed by the perfect and asymmetric palindromes may be due to the inability of DNA helicases to unwind stable stem-loop structures. In this case, the large loop formed by the asymmetric palindrome in DRL180 may destabilise the terminal base pairs in the hairpin stem. While this may simply be an effect of loop folding on helix stability, it is interesting to speculate whether the interaction of loop nucleotides with SSB (single-stranded DNA binding protein) may facilitate hairpin stem unwinding. The stoichiometry of 65 nucleotides per SSB tetramer under physiological conditions (Meyer & Laine, 1990) is in reasonable accord with the minimum length of central asymmetry (57 bp) required for complete alleviation of palindrome-mediated inviability (Warren & Green, 1985).

The significantly greater viability of DRL180 may therefore be due to the higher efficiency of replication of hairpin structures with large loops. When hairpins are formed by processes involving DNA unwinding such as transcription, the result is a cruciform structure which will inhibit Dam methylation at the palindrome centre. The methylation data for DRL180 (Chapter 9) indicates that hairpin structures of this kind are formed frequently. Perfect palindromes are similarly capable of forming hairpins on both strands (yielding cruciforms), but may also undergo cruciform extrusion in negatively-supercoiled double-stranded molecules as shown in Fig. 1.10. It is therefore puzzling that one of the perfect palindromes (DRL176) should exhibit less methylation inhibition (inaccessible for ~36% of λ lifecycle) than the asymmetric palindrome (inaccessible for ~55% of λ lifecycle).

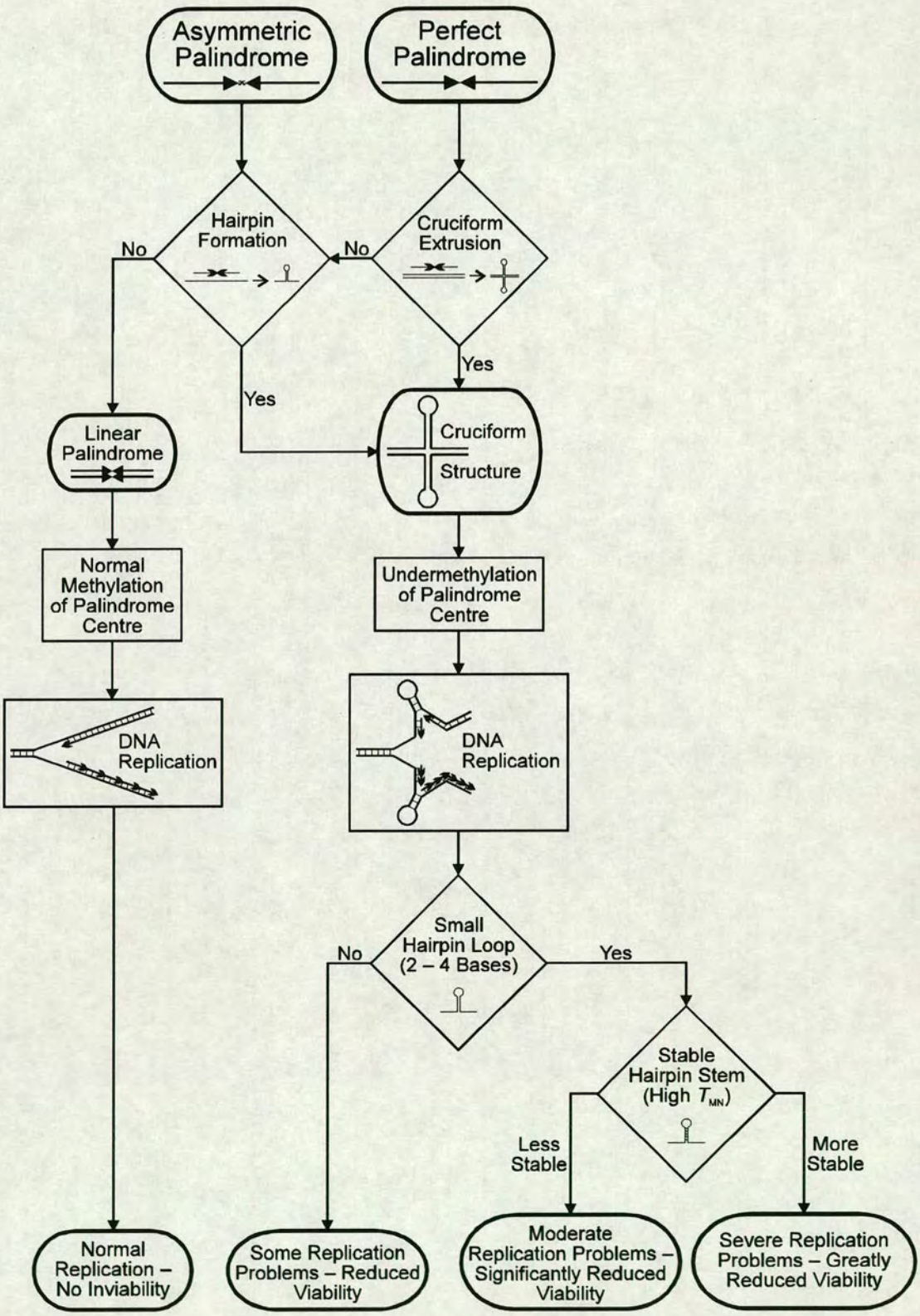
The equivalent levels of methylation inhibition seen at the centres of perfect and asymmetric palindromes may be due to an underestimation of cruciform formation by perfect palindromes *in vivo*. The results presented in Chapter 10 indicate that perfect palindromes replicate poorly by comparison with asymmetric palindromes. If this is due to the arrest of replication fork progression by stable stem-loop structures, then DNA molecules that have undergone hairpin formation or cruciform extrusion will not be replicated. Consequently, the frequency of cruciform formation *in vivo* will be underestimated by the Dam methylation assay in the case of the perfect palindromes (see Conclusion to Chapter 9). As the asymmetric palindrome does not lead to comparable inviability, the methylation inhibition at its centre will reflect more accurately the incidence of hairpin or cruciform formation *in vivo*. Only if it were possible to alleviate completely the inviability associated with long (perfect) palindromes, would the undermethylation observed at these palindrome centres be commensurate with cruciform formation *in vivo*.

In summary, the hypothesis presented here proposes that perfect palindromes undergo cruciform extrusion *in vivo* by a centre-dependent pathway. This structural transition is kinetically forbidden to asymmetric palindromes. However, both perfect and asymmetric palindromes may form hairpins when present in single-stranded regions, thereby yielding cruciform structures. These unusual DNA secondary structures inhibit Dam methylation at the palindrome centre. They may also compromise DNA replication, thereby leading to replicon inviability. The severity of replication problems depends primarily on the size of the hairpin loop, and secondarily on the thermal stability of the hairpin stem. Replication inhibition by stable hairpins with small loops may also bias the methylation assay by favouring the detection of unextruded perfect palindromes, which are replicated and methylated efficiently.

This hypothesis proposes that palindrome-mediated inviability is not a direct consequence of cruciform extrusion. It is therefore in keeping with previous studies showing the lack of a clear relationship between the physical and genetic properties of long DNA palindromes (Warren & Green, 1985). The hypothesis is illustrated schematically by the flow-chart algorithm in Fig. 12.1.

Figure 12.1

Flow-chart algorithm for palindrome-mediated inviability and methylation inhibition.



BIBLIOGRAPHY

Albertini, A.M., Hofer, M., Calos, M.P. & Miller, J.H. (1982) On the formation of spontaneous deletions: The importance of short sequence homologies in the generation of large deletions. *Cell* **29:** 319–328

Allgood, N.D. & Silhavy, T.J. (1991) *Escherichia coli xonA* (*sbcB*) mutants enhance illegitimate recombination. *Genetics* **127:** 671–680

Altenbuchner, J., Viell, P. & Pelletier, I. (1992) Positive selection vectors based on palindromic DNA sequences. *In:* Wu, R. (ed.) Recombinant DNA Part G. pp. 457–466, *Methods in Enzymology* Vol. 216 (Academic Press Inc., San Diego)

Astell, C.R., Chow, M.B. & Ward, D.C. (1985) Sequence analysis of the termini or virion and replicative forms of minute virus of mice DNA suggests a modified rolling hairpin model for autonomous parvovirus DNA replication. *J. Virol.* **54:** 171–177

Astell, C.R., Thomson, M., Chow, M.B. & Ward, D.C. (1983) Structure and replication of minute virus of mice DNA. *Cold Spring Harbor Symp. Quant. Biol.* **47:** 751–762

Bachmann, B.J. (1972) Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol. Rev.* **36:** 525–557

Backman, K., Betlach, M., Boyer, H.W. & Yanofsky, S. (1978) Genetic and physical studies on the replication of ColE1-type plasmids. *Cold Spring Harbor Symp. Quant. Biol.* **43:** 69–76

Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Séguin, C., Tuffnell, P.S. & Barrell, B.G. (1984) DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature (Lond.)* **310:** 207–211

Bagga, R., Ramesh, N. & Brahmachari, S.K. (1990) Supercoil-induced unusual DNA structures as transcriptional block. *Nucleic Acids Res.* **18:** 3363–3369

Baroudy, B.M., Venkatesan, S. & Moss, B. (1982) Structure and replication of vaccinia virus telomeres. *Cold Spring Harbor Symp. Quant. Biol.* **47:** 723–729

Barras, F. & Marinus, M.G. (1989) The great GATC: DNA methylation in *E. coli*. *Trends Genet.* **5:** 139–143

Bassett, C.L. & Kushner, S.R. (1984) Exonucleases I, III, and V are required for stability of ColE1-related plasmids in *Escherichia coli*. *J. Bacteriol.* **157:** 661–664

Bauer, W. & Vinograd, J. (1968) The interaction of closed circular DNA with intercalative dyes: I. The superhelix density of SV40 DNA in the presence and absence of dye. *J. Mol. Biol.* **33:** 141–171

Bauer, W. & Vinograd, J. (1970) Interaction of closed circular DNA with intercalative dyes: II. The free energy of superhelix formation in SV40 DNA. *J. Mol. Biol.* **47:** 419–435

Behnke, D., Malke, H., Hartmann, M. & Walter, F. (1979) Post-transformational rearrangement of an *in vitro* reconstructed group-A streptococcal erythromycin resistance plasmid. *Plasmid* **2:** 605–616

Bell, L. & Byers, B. (1979) Occurrence of crossed strand-exchange forms in yeast DNA during meiosis. *Proc. Natl. Acad. Sci. USA* **76:** 3445–3449

Bell, L.R. & Byers, B. (1983) Homologous association of chromosomal DNA during yeast meiosis. *Cold Spring Harbor Symp. Quant. Biol.* **47:** 829–840

Bell, D., Sabloff, M., Zannis-Hadjopoulos, M. & Price, G. (1991) Anti-cruciform DNA affinity purification of active mammalian origins of replication. *Biochim. Biophys. Acta* **1089:** 299–308

Benbow, R.M., Zuccarelli, A.J. & Sinsheimer, R.L. (1975) Recombinant DNA molecules of bacteriophage ϕX174. *Proc. Natl. Acad. Sci. USA* **72:** 235–239

Benham, C.J. (1982) Stable cruciform formation at inverted repeat sequences in supercoiled DNA. *Biopolymers* **21:** 679–696

Benham, C.J. (1992) Energetics of the strand separation transition in superhelical DNA. *J. Mol. Biol.* **225:** 835–847

Berg, D.E. (1989) Transposon Tn*5*. *In:* Berg, D.E. & Howe, M.M. (ed.) Mobile DNA. pp. 185–210 (American Society for Microbiology, Washington, D.C.)

Berg, D.E., Egner, C., Hirschel, B.J., Howard, J., Johnsrud, L., Jorgensen, R.A. & Tlsty, T.D. (1981) Insertion, excision and inversion of Tn*5*. *Cold Spring Harbor Symp. Quant. Biol.* **45:** 115–123

Berg, D.E., Egner, C. & Lowe, J.B. (1983) Mechanism of F factor-enhanced excision of transposon Tn*5*. *Gene (Amst.)* **22:** 1–7

Bergsma, D.J., Olive, D.M., Hartzell, S.W., Byrne, B.J. & Subramanian, K.N. (1982) Cyclization of linear chimeric plasmids *in vivo* by a novel end-to-end joining reaction or by intramolecular recombination: One of the products contains a 147-bp perfect palindrome stable in *Escherichia coli*. *Gene (Amst.)* **20:** 157–167

Bernardi, G. (1962) Chromatography of denatured deoxyribonucleic acid on calcium phosphate. *Biochem. J.* **83:** 32p–33p

Bertrand, K., Korn, L., Lee, F., Platt, T., Squires, C.L., Squires, C. & Yanofsky, C. (1975) New features of the regulation of the tryptophan operon: A new type of regulatory site has been studied. *Science (Washington DC)* **189:** 22–26

Betz, J.L. & Sadler, J.R. (1981) Variants of a cloned synthetic lactose operator: I. A palindromic dimer lactose operator derived from one strand of the cloned 40-base pair operator. *Gene (Amst.)* **13:** 1–12

Bhattacharyya, A., Murchie, A.I.H., von Kitzing, E., Diekmann, S., Kemper, B. & Lilley, D.M.J. (1991) Model for the interaction of DNA junctions and resolving enzymes. *J. Mol. Biol.* **221:** 1191–1207

Bianchi, M.E. (1988) Interaction of a protein from rat liver nuclei with cruciform DNA. *EMBO (Eur. Mol. Biol. Organ.) J.* **7:** 843–849

Bianchi, M.E., Beltrame, M. & Paonessa, G. (1989) Specific recognition of cruciform DNA by nuclear protein HMG1. *Science (Washington DC)* **243:** 1056–1059

Bianchi, M.E., Falciola, L., Ferrari, S. & Lilley, D.M.J. (1992) The DNA binding site of HMG1 protein is composed of two similar segments (HMG boxes), both of which have counterparts in other eukaryotic regulatory proteins. *EMBO (Eur. Mol. Biol. Organ.) J.* **11:** 1055–1063

Bingham, P.M. & Zachar, Z. (1989) Reterotransposons and the FB transposon from *Drosophila melanogaster*. *In:* Berg, D.E. & Howe, M.M. (ed.) Mobile DNA. pp. 485–502 (American Society for Microbiology, Washington, D.C.)

Birchmeier, C., Folk, W. & Birnstiel, M.L. (1983) The terminal RNA stem-loop structure and 80 bp of spacer DNA are required for the formation of 3' termini of sea urchin H2A mRNA. *Cell* **35:** 433–440

Birnboim, H.C. & Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7:** 1513–1523

Birnstiel, M.L., Busslinger, M. & Strub, K. (1985) Transcription termination and 3' processing: The end is in site! *Cell* **41:** 349–359

Blaho, J.A. & Wells, R.D. (1987) Left-handed Z-DNA binding by the recA protein of *Escherichia coli*. *J. Biol. Chem.* **262:** 6082–6088

Bliska, J.B. & Cozzarelli, N.R. (1987) Use of site-specific recombination as a probe of DNA structure and metabolism *in vivo*. *J. Mol. Biol.* **194:** 205–218

Blommers, M.J.J., Van de Ven, F.J.M., Van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1991) The three-dimensional structure of a DNA hairpin in solution. *Eur. J. Biochem.* **201:** 33–51

Blommers, M.J.J., Walters, J.A.L.I., Haasnoot, C.A.G., Aelen, J.M.A., Van der Marel, G.A., van Boom, J.H. & Hilbers, C.W. (1989) Effects of base sequence on the loop folding in DNA hairpins. *Biochemistry* **28:** 7491–7498

Boissy, R. & Astell, C.R. (1985) An *Escherichia coli recBCsbcBrecF* host permits the deletion-resistant propagation of plasmid clones containing the 5'-terminal palindrome of minute virus of mice. *Gene (Amst.)* **35:** 179–185

Bolivar, F., Betlach, M.C., Heyneker, H.L., Shine, J., Rodriguez, R.L. & Boyer, H.W. (1977a) Origin of replication of pBR345 plasmid DNA. *Proc. Natl. Acad. Sci. USA* **74:** 5265–5269

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. & Falkow, S. (1977b) Construction and characterization of new cloning vehicles: II. A multipurpose cloning system. *Gene (Amst.)* **2:** 95–113

Borst, P., Overdulve, J.P., Weijers, P.J., Fase-Fowler, F. & Van den Berg, M. (1984) DNA circles with cruciforms from *Isospora* (*Toxoplasma*) *gondii*. *Biochim. Biophys. Acta* **781:** 100–111

Bouffler, S., Silver, A. & Cox, R. (1993) The role of DNA repeats and associated secondary structures in genomic instability and neoplasia. *BioEssays* **15:** 409–412

Boulard, Y., Gabarro-Arpa, J., Cognet, J.A.H., Le Bret, M., Guy, A., Téoule, R., Guschlbauer, W. & Fazakerley, G.V. (1991) The solution structure of a DNA hairpin containing a loop of three thymidines determined by nuclear magnetic resonance and molecular mechanics. *Nucleic Acids Res.* **19:** 5159–5167

Boulikas, T. (1992) Evolutionary consequences of nonrandom damage and repair of chromatin domains. *J. Mol. Evol.* **35:** 156–180

Bowater, R., Aboul-ela, F. & Lilley, D.M.J. (1991) Large-scale stable opening of supercoiled DNA in response to temperature and supercoiling in (A+T)-rich regions that promote low-salt cruciform extrusion. *Biochemistry* **30:** 11495–11506

Bowater, R., Aboul-ela, F. & Lilley, D.M.J. (1992) Two-dimensional gel electrophoresis of circular DNA topoisomers. *In:* Lilley, D.M.J. & Dahlberg, J.E. (ed.) DNA Structures Part B: Chemical and Electrophoretic Analysis of DNA. pp. 105–120, *Methods in Enzymology* Vol. 212 (Academic Press Inc., San Diego)

Boyer, H.W. & Roulland-Dussoix, D. (1969) A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J. Mol. Biol.* **41:** 459–472

Brahmachari, S.K., Shouche, Y.S., Cantor, C.R. & McClelland, M. (1987) Sequences that adopt non-B-DNA conformation in form V DNA as probed by enzymic methylation. *J. Mol. Biol.* **193:** 201–211

Broker, T.R. & Lehman, I.R. (1971) Branched DNA molecules: Intermediates in T4 recombination. *J. Mol. Biol.* **60:** 131–149

Broyles, S.S. & Pettijohn, D.E. (1986) Interaction of the *Escherichia coli* HU protein with DNA: Evidence for formation of nucleosome-like structures with altered DNA helical pitch. *J. Mol. Biol.* **187:** 47–60

Brunier, D., Michel, B. & Ehrlich, S.D. (1988) Copy choice illegitimate DNA recombination. *Cell* **52:** 883–892

Brunier, D., Peeters, B.P.H., Bron, S. & Ehrlich, S.D. (1989) Breakage-reunion and copy choice mechanisms of recombination between short homologous sequences. *EMBO (Eur. Mol. Biol. Organ.) J.* **8:** 3127–3133

Burland, V., Plunkett, G., Daniels, D.L. & Blattner, F.R. (1993) DNA sequence and analysis of 136 kilobases of the *Escherichia coli* genome: Organizational symmetry around the origin of replication. *Genomics* **16:** 551–561

Burt, D.W. & Brammar, W.J. (1982a) Transcriptional termination sites in the *b2* region of bacteriophage lambda that are unresponsive to antitermination. *Mol. & Gen. Genet.* **185:** 462–467

Burt, D.W. & Brammar, W.J. (1982b) The *cis*-specificity of the *Q*-gene product of bacteriophage lambda. *Mol. & Gen. Genet.* **185:** 468–472

Campbell, A. (1965) The steric effect in lysogenization by bacteriophage lambda: I. Lysogenization of a partially diploid strain of *Escherichia coli* K12. *Virology* **27:** 329–339

Campbell, J.L. & Kleckner, N. (1990) *E. coli oriC* and the *dnaA* gene promoter are sequestered from *dam* methyltransferase following the passage of the chromosomal replication fork. *Cell* **62:** 967–979

Casadaban, M.J. & Cohen, S.N. (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J. Mol. Biol.* **138:** 179–207

Cavalier-Smith, T. (1974) Palindromic base sequences and replication of eukaryote chromosome ends. *Nature (Lond.)* **250:** 467–470

Chalker, A.F. (1990) *sbcC* and palindrome-mediated inviability in *Escherichia coli*. *Ph.D. Thesis*, 174 pp. (University of Edinburgh)

Chalker, A.F., Leach, D.R.F. & Lloyd, R.G. (1988) *Escherichia coli sbcC* mutants permit stable propagation of DNA replicons containing a long palindrome. *Gene (Amst.)* **71:** 201–205

Chalker, A.F., Okely, E.A., Davison, A. & Leach, D.R.F. (1993) The effects of central asymmetry on the propagation of palindromic DNA in bacteriophage λ are consistent with cruciform extrusion *in vivo*. *Genetics* **133:** 143–148

Chen, S., Heffron, F., Leupin, W. & Chazin, W.J. (1991) Two-dimensional ^{1}H NMR studies of synthetic immobile Holliday junctions. *Biochemistry* **30:** 766–771

Chow, L.T., Davidson, N. & Berg, D. (1974) Electron microscope study of the structures of λ*dv* DNAs. *J. Mol. Biol.* **86:** 69–89

Churchill, M.E.A., Tullius, T.D., Kallenbach, N.R. & Seeman, N.C. (1988) A Holliday recombination intermediate is twofold symmetric. *Proc. Natl. Acad. Sci. USA* **85:** 4653–4656

Clark, A.J. & Low, K.B. (1988) Pathways and systems of homologous recombination in *Escherichia coli*. *In:* Low, K.B. (ed.) The Recombination of Genetic Material. pp. 155–215 (Academic Press, Inc., San Diego, California)

Clark, A.J. & Margulies, A.D. (1965) Isolation and characterization of recombination-deficient mutants of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **53:** 451–459

Clegg, R.M., Murchie, A.I.H., Zechel, A., Carlberg, C., Diekmann, S. & Lilley, D.M.J. (1992) Fluorescence resonance energy transfer analysis of the structure of the four-way DNA junction. *Biochemistry* **31:** 4846–4856

Cohen, A. & Clark, A.J. (1986) Synthesis of linear plasmid multimers in *Escherichia coli* K-12. *J. Bacteriol.* **167:** 327–335

Collins, J. (1981) Instability of palindromic DNA in *Escherichia coli*. *Cold Spring Harbor Symp. Quant. Biol.* **45:** 409–416

Collins, J. & Hohn, B. (1978) Cosmids: A type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage λ heads. *Proc. Natl. Acad. Sci. USA* **75:** 4242–4246

Collins, M. & Myers, R.M. (1987) Alterations in DNA helix stability due to base modifications can be evaluated using denaturing gradient gel electrophoresis. *J. Mol. Biol.* **198:** 737–744

Collins, J., Volckaert, G. & Nevers, P. (1982) Precise and nearly-precise excision of the symmetrical inverted repeats of Tn*5*; common features of *recA*-independent deletion events in *Escherichia coli*. *Gene (Amst.)* **19:** 139–146

Connolly, B., Parsons, C.A., Benson, F.E., Dunderdale, H.J., Sharples, G.J., Lloyd, R.G. & West, S.C. (1991) Resolution of Holliday junctions *in vitro* requires the *Escherichia coli ruvC* gene product. *Proc. Natl. Acad. Sci. USA* **88:** 6063–6067

Cooper, J.P. & Hagerman, P.J. (1987) Gel electrophoretic analysis of the geometry of a DNA four-way junction. *J. Mol. Biol.* **198:** 711–719

Cooper, J.P. & Hagerman, P.J. (1989) Geometry of a branched DNA structure in solution. *Proc. Natl. Acad. Sci. USA* **86:** 7336–7340

Cotmore, S.F., Nüesch, J.P.F. & Tattersall, P. (1993) Asymmetric resolution of a parvovirus palindrome *in vitro*. *J. Virol.* **67:** 1579–1589

Courey, A.J. & Wang, J.C. (1983) Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions. *Cell* **33:** 817–829

Courey, A.J. & Wang, J.C. (1988) Influence of DNA sequence and supercoiling on the process of cruciform extrusion. *J. Mol. Biol.* **202:** 35–43

Crothers, D.M. & Drak, J. (1992) Global features of DNA structure by comparative gel electrophoresis. *In:* Lilley, D.M.J. & Dahlberg, J.E. (ed.) DNA Structures Part B: Chemical and Electrophoretic Analysis of DNA. pp. 46–71, *Methods in Enzymology* Vol. 212 (Academic Press Inc., San Diego)

Crouse, G.F. (1985) Plasmids supplying the Q-*qut*-controlled Gam function permit growth of λ *red*$^-$ *gam*$^-$ (Fec$^-$) bacteriophages on *recA*$^-$ hosts. *Gene (Amst.)* **40:** 151–155

Cunningham, R.P. & Berger, H. (1977) Mutations affecting genetic recombination in bacteriophage T4D. *Virology* **80:** 67–82

Dabert, P., Ehrlich, S.D. & Gruss, A. (1992) χ sequence protects against RecBCD degradation of DNA *in vivo*. *Proc. Natl. Acad. Sci. USA* **89:** 12073–12077

DasGupta, U., Weston-Hafer, K. & Berg, D.E. (1987) Local DNA sequence control of deletion formation in *Escherichia coli* plasmid pBR322. *Genetics* **115:** 41–49

DasGupta, C., Wu, A.M., Kahn, R., Cunningham, R.P. & Radding, C.M. (1981) Concerted strand exchange and formation of Holliday structures by *E. coli* RecA protein. *Cell* **25:** 507–516

Dayn, A., Malkhosyan, S., Duzhy, D., Lyamichev, V., Panchenko, Y. & Mirkin, S. (1991) Formation of $(dA\text{-}dT)_n$ cruciforms in *Escherichia coli* cells under different environmental conditions. *J. Bacteriol.* **173:** 2658–2664

Davison, A. & Leach, D.R.F. *Genetics* (in the press)

Dayn, A., Malkhosyan, S. & Mirkin, S.M. (1992) Transcriptionally driven cruciform formation *in vivo*. *Nucleic Acids Res.* **20:** 5991–5997

de Boer, J.G. & Ripley, L.S. (1984) Demonstration of the production of frameshift and base-substitution mutations by quasipalindromic DNA sequences. *Proc. Natl. Acad. Sci. USA* **81:** 5528–5531

de Massy, B., Studier, F.W., Dorgai, L., Appelbaum, E. & Weisberg, R.A. (1985) Enzymes and sites of genetic recombination: Studies with gene-*3* endonuclease of phage T7 and with site-affinity mutants of phage λ. *Cold Spring Harbor Symp. Quant. Biol.* **49:** 715–726

de Massy, B., Weisberg, R.A. & Studier, F.W. (1987) Gene *3* endonuclease of bacteriophage T7 resolves conformationally branched structures in double-stranded DNA. *J. Mol. Biol.* **193:** 359–376

del Olmo, M.L. & Pérez-Ortín, J.E. (1993) A natural A/T-rich sequence from the yeast *FBP1* gene exists as a cruciform in *Escherichia coli* cells. *Plasmid* **29:** 222–232

Depew, R.E. & Wang, J.C. (1975) Conformational fluctuations of DNA helix. *Proc. Natl. Acad. Sci. USA* **72:** 4275–4279

Diekmann, S. & Lilley, D.M.J. (1987) The anomalous gel migration of a stable cruciform: Temperature and salt dependence, and some comparisons with curved DNA. *Nucleic Acids Res.* **15:** 5765–5774

Dimri, G.P. & Das, H.K. (1990) Cloning and sequence analysis of *gyrA* gene of *Klebsiella pneumoniae*. *Nucleic Acids Res.* **18:** 151–156

Dingwall, C., Lomonossoff, G.P. & Laskey, R.A. (1981) High sequence specificity of micrococcal nuclease. *Nucleic Acids Res.* **9:** 2659–2673

Doherty, J.P., Lindeman, R., Trent, R.J., Graham, M.W. & Woodcock, D.M. (1993) *Escherichia coli* host strains SURE™ and SRB fail to preserve a palindrome cloned in lambda phage: Improved alternate host strains. *Gene (Amst.)* **124:** 29–35

Doniger, J., Warner, R.C. & Tessma, I. (1973) Role of circular dimer DNA in the primary recombination mechanism of bacteriophage S13. *Nature New Biol.* **242:** 9–12

Donlon, T.A., Lalande, M., Wyman, A., Bruns, G. & Latt, S.A. (1986) Isolation of molecular probes associated with chromosome 15 instability in the Prader-Willi syndrome. *Proc. Natl. Acad. Sci. USA* **83:** 4408–4412

Drew, H.R., Weeks, J.R. & Travers, A.A. (1985) Negative supercoiling induces spontaneous unwinding of a bacterial promoter. *EMBO (Eur. Mol. Biol. Organ.) J.* **4:** 1025–1032

Drlica, K. (1990) Bacterial topoisomerases and the control of DNA supercoiling. *Trends Genet.* **6:** 433–437

Drlica, K. & Rouviere-Yaniv, J. (1987) Histonelike proteins of bacteria. *Microbiol. Rev.* **51:** 301–319

Drolet, M. & Lau, P.C.K. (1992) Mobilization protein-DNA binding and divergent transcription at the transfer origin of the *Thiobacillus ferrooxidans* pTF1 plasmid. *Mol. Microbiol.* **6:** 1061–1071

Duckett, D.R. & Lilley, D.M.J. (1991) Effects of base mismatches on the structure of the four-way DNA junction. *J. Mol. Biol.* **221:** 147–161

Duckett, D.R., Murchie, A.I.H., Bhattacharyya, A., Clegg, R.M., Diekmann, S., von Kitzing, G. & Lilley, D.M.J. (1992) The structure of DNA junctions and their interaction with enzymes. *Eur. J. Biochem.* **207:** 285–295

Duckett, D.R., Murchie, A.I.H., Diekmann, S., von Kitzing, E., Kemper, B. & Lilley, D.M.J. (1988) The structure of the Holliday junction, and its resolution. *Cell* **55:** 79–89

Duckett, D.R., Murchie, A.I.H. & Lilley, D.M.J. (1990) The role of metal ions in the conformation of the four-way DNA junction. *EMBO (Eur. Mol. Biol. Organ.) J.* **9:** 583–590

Dunderdale, H.J., Benson, F.E., Parsons, C.A., Sharples, G.J., Lloyd, R.G. & West, S.C. (1991) Formation and resolution of recombination intermediates by *E. coli* RecA and RuvC proteins. *Nature (Lond.)* **354:** 506–510

Dykes, G., Bambara, R., Marian, K. & Wu, R. (1975) On the statistical significance of primary structural features found in DNA-protein interaction sites. *Nucleic Acids Res.* **2:** 327–345

Egner, C. & Berg, D.E. (1981) Excision of transposon Tn*5* is dependent on the inverted repeats but not on the transposase function of Tn*5*. *Proc. Natl. Acad. Sci. USA* **78:** 459–463

Ehrlich, S.D. (1989) Illegitimate recombination in bacteria. *In:* Berg, D.E. & Howe, M.M. (ed.) Mobile DNA. pp. 799–832 (American Society for Microbiology, Washington, D.C.)

Elborough, K.M. & West, S.C. (1988) Specific binding of cruciform DNA structures by a protein from human extracts. *Nucleic Acids Res.* **16:** 3603–3616

Elborough, K.M. & West, S.C. (1990) Resolution of synthetic Holliday junctions in DNA by an endonuclease activity from calf thymus. *EMBO (Eur. Mol. Biol. Organ.) J.* **9:** 2931–2936

Elhai, J. & Wolk, C.P. (1988) A versatile class of positive-selection vectors based on the nonviability of palindrome-containing plasmids that allows cloning into long polylinkers. *Gene (Amst.)* **68:** 119–138

Engberg, J., Andersson, P., Leick, V. & Collins, J. (1976) Free ribosomal DNA molecules from *Tetrahymena pyriformis* GL are giant palindromes. *J. Mol. Biol.* **104:** 455–470

Enquist, L.W. & Skalka, A. (1973) Replication of bacteriophage λ DNA dependent on the function of host and viral genes. *J. Mol. Biol.* **75:** 185–212

Erie, D.A., Suri, A.K., Breslauer, K.J., Jones, R.A. & Olson, W.K. (1993) Theoretical predictions of DNA hairpin loop conformations: Correlations with thermodynamic and spectroscopic data. *Biochemistry* **32:** 436–454

Evans, D.H. & Kolodner, R. (1988) Effect of DNA structure and nucleotide sequence on Holliday junction resolution by a *Saccharomyces cerevisiae* endonuclease. *J. Mol. Biol.* **201:** 69–80

Ezekiel, U.R. & Zassenhaus, H.P. (1993) Localization of a cruciform cutting endonuclease to yeast mitochondria. *Mol. & Gen. Genet.* **240:** 414–418

Farabaugh, P.J., Schmeissner, U., Hofer, M. & Miller, J.H. (1978) Genetic studies of the *lac* repressor: VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. *J. Mol. Biol.* **126:** 847–863

Farnham, P.J. & Platt, T. (1981) Rho-independent termination: Dyad symmetry in DNA causes RNA polymerase to pause during transcription *in vitro*. *Nucleic Acids Res.* **9:** 563–577

Ferrari, S., Harley, V.R., Pontiggia, A., Goodfellow, P.N., Lovell-Badge, R. & Bianchi, M.E. (1992) SRY, like HMG1, recognizes sharp angles in DNA. *EMBO (Eur. Mol. Biol. Organ.) J.* **11:** 4497–4506

Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G. & Ysebaert, M. (1978) Complete nucleotide sequence of SV40 DNA. *Nature (Lond.)* **273:** 113–120

Foster, T.J., Lundblad, V., Hanley-Way, S., Halling, S.M. & Kleckner, N. (1981) Three Tn*10*-associated excision events: Relationship to transposition and role of direct and inverted repeats. *Cell* **23:** 215–227

Fowler, R.F. & Skinner, D.M. (1985) Cryptic satellites rich in inverted repeats comprise 30% of the genome of a hermit crab. *J. Biol. Chem.* **260:** 1296–1303

Frank-Kamenetskii, M.D. (1990) DNA supercoiling and unusual structures. *In:* Cozzarelli, N.R. & Wang, J.C. (ed.) DNA Topology and Its Biological Effects. pp. 185–215, *Cold Spring Harbor Monograph Series* Vol. 20 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York)

Frank-Kamenetskii, M.D. & Vologodskii, A.V. (1984) Thermodynamics of the B-Z transition in superhelical DNA. *Nature (Lond.)* **307:** 481–482

Franklin, N.C. (1967) Extraordinary recombinational events in *Escherichia coli*. Their independence of the *rec*$^+$ function. *Genetics* **55:** 699–707

Franklin, N.C. (1989) A plasmid to visualize and assay termination and antitermination of transcription in *Escherichia coli*. *Plasmid* **21:** 31–42

Frappier, L., Price, G.B., Martin, R.G. & Zannis-Hadjopoulos, M. (1987) Monoclonal antibodies to cruciform DNA structures. *J. Mol. Biol.* **193:** 751–758

Frappier, L., Price, G.B., Martin, R.G. & Zannis-Hadjopoulos, M. (1989) Characterization of the binding specificity of two anticruciform DNA monoclonal antibodies. *J. Biol. Chem.* **264:** 334–341

Friedberg, E.C. (1985) DNA repair. 614 pp. (W. H. Freeman and Company, New York)

Friedman, D.I. & Gottesman, M.E. (1983) Lytic mode of lambda development. *In:* Hendrix, R.W., Roberts, J.W., Stahl, F.W. & Weisberg, R.A. (ed.) Lambda II. pp. 21–51, *Cold Spring Harbor Monograph Series* Vol. 13 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York)

Frisque, R.J. (1983) Nucleotide sequence of the region encompassing the JC virus origin of DNA replication. *J. Virol.* **46:** 170–176

Furlong, J.C. & Lilley, D.M.J. (1986) Highly selective chemical modification of cruciform loops by diethyl pyrocarbonate. *Nucleic Acids Res.* **14:** 3995–4007

Galas, D.J. & Chandler, M. (1989) Bacterial insertion sequences. *In:* Berg, D.E. & Howe, M.M. (ed.) Mobile DNA. pp. 109–162 (American Society for Microbiology, Washington, D.C.)

Gelinas, R.E., Myers, P.A. & Roberts, R.J. (1977) Two sequence-specific endonucleases from *Moraxella bovis*. *J. Mol. Biol.* **114:** 169–179

Gellert, M., Mizuuchi, K., O'Dea, M.H., Ohmori, H. & Tomizawa, J. (1979) DNA gyrase and DNA supercoiling. *Cold Spring Harbor Symp. Quant. Biol.* **43:** 35–40

Gellert, M., O'Dea, M.H. & Mizuuchi, K. (1983) Slow cruciform transitions in palindromic DNA. *Proc. Natl. Acad. Sci. USA* **80:** 5545–5549

Germann, M.W., Kalisch, B.W., Lundberg, P., Vogel, H.J. & Van der Sande, J.H. (1990) Perturbation of DNA hairpins containing the *Eco*RI recognition site by hairpin loops of varying size and composition: Physical (NMR and UV) and enzymatic (*Eco*RI) studies. *Nucleic Acids Res.* **18:** 1489–1498

Gibson, F.P., Leach, D.R.F. & Lloyd, R.G. (1992) Identification of *sbcD* mutations as cosuppressors of *recBC* that allow propagation of DNA palindromes in *Escherichia coli* K-12. *J. Bacteriol.* **174:** 1222–1228

Gierer, A. (1966) Model for DNA and protein interactions and the function of the operator. *Nature (Lond.)* **212:** 1480–1481

Gilson, E., Bachellier, S., Perrin, S., Perrin, D., Grimont, P.A.D., Grimont, F. & Hofnung, M. (1990a) Palindromic unit highly repetitive DNA sequences exhibit species specificity within *Enterobacteriaceae*. *Res. Microbiol.* **141:** 1103–1116

Gilson, E., Clément, J., Brutlag, D. & Hofnung, M. (1984) A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. *EMBO (Eur. Mol. Biol. Organ.) J.* **3:** 1417–1421

Gilson, E., Clément, J.M., Perrin, D. & Hofnung, M. (1987) Palindromic units: A case of highly repetitive DNA sequences in bacteria. *Trends Genet.* **3:** 226–230

Gilson, E., Perrin, D. & Hofnung, M. (1990b) DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic unit highly repetitive DNA: Implications for bacterial chromosome organisation. *Nucleic Acids Res.* **18:** 3941–3952

Gilson, E., Saurin, W., Perrin, D., Bachellier, S. & Hofnung, M. (1991) Palindromic units are part of a new bacterial interspersed mosaic element (BIME). *Nucleic Acids Res.* **19:** 1375–1383

Glasgow, A.C., Hughes, K.T. & Simon, M.I. (1989) Bacterial DNA inversion systems. *In:* Berg, D.E. & Howe, M.M. (ed.) Mobile DNA. pp. 637–659 (American Society for Microbiology, Washington, D.C.)

Glickman, B.W. & Ripley, L.S. (1984) Structural intermediates of deletion mutagenesis: A role for palindromic DNA. *Proc. Natl. Acad. Sci. USA* **81:** 512–516

Glucksmann, M.A., Markiewicz, P., Malone, C. & Rothman-Denes, L.B. (1992) Specific sequences and a hairpin structure in the template strand are required for N4 virion RNA polymerase promoter recognition. *Cell* **70:** 491–500

Gorbalenya, A.E. & Koonin, E.V. (1990) Superfamily of UvrA-related NTP-binding proteins: Implications for rational classification of recombination/repair systems. *J. Mol. Biol.* **213:** 583–591

Gordenin, D.A., Lobachev, K.S., Degtyareva, N.P., Malkova, A.L., Perkins, E. & Resnick, M.A. (1993) Inverted DNA repeats: A source of eukaryotic genomic instability. *Mol. Cell. Biol.* **13:** 5315–5322

Gordenin, D.A., Malkova, A.L., Peterzen, A., Kulikov, V.N., Pavlov, Y.I., Perkins, E. & Resnick, M.A. (1992) Transposon Tn*5* excision in yeast: Influence of DNA polymerases α, δ, and ε and repair genes. *Proc. Natl. Acad. Sci. USA* **89:** 3785–378

Gotoh, O. & Tagashira, Y. (1981) Stabilities of nearest-neighbour doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles. *Biopolymers* **20:** 1033–1042

Gottesman, M.E., Adhya, S. & Das, A. (1980) Transcription antitermination by bacteriophage lambda *N* gene product. *J. Mol. Biol.* **140:** 57–75

Gough, G.W. & Lilley, D.M.J. (1985) DNA bending induced by cruciform formation. *Nature (Lond.)* **313:** 154–156

Gough, G.W., Sullivan, K.M. & Lilley, D.M.J. (1986) The structure of cruciforms in supercoiled DNA: Probing the single-stranded character of nucleotide bases with bisulphite. *EMBO (Eur. Mol. Biol. Organ.) J.* **5:** 191–196

Gram, H. & Rüger, W. (1985) Genes *55*, *αgt*, *47* and *46* of bacteriophage T4: The genomic organisation as deduced by sequence analysis. *EMBO (Eur. Mol. Biol. Organ.) J.* **4:** 257–264

Greaves, D.R., Patient, R.K. & Lilley, D.M.J. (1985) Facile cruciform formation by an $(A\text{-}T)_{34}$ sequence from a Xenopus globin gene. *J. Mol. Biol.* **185:** 461–478

Greenstein, M. & Skalka, A. (1975) Replication of bacteriophage lambda DNA: *In vivo* studies of the interaction between the viral *gamma* protein and the host *recBC* DNAase. *J. Mol. Biol.* **97:** 543–559

Grindley, N.D.F. & Reed, R.R. (1985) Transpositional recombination in prokaryotes. *Annu. Rev. Biochem.* **54:** 863–896

Hagan, C.E. & Warren, G.J. (1982) Lethality of palindromic DNA and its use in selection of recombinant plasmids. *Gene (Amst.)* **19:** 147–151

Hagan, C.E. & Warren, G.J. (1983) Viability of palindromic DNA is restored by deletions occurring at low but variable frequency in plasmids of *Escherichia coli*. *Gene (Amst.)* **24:** 317–326

Haniford, D.B. & Pulleyblank, D.E. (1983) Facile transition of poly[d(TG)·d(CA)] into a left-handed helix in physiological conditions. *Nature (Lond.)* **302:** 632–634

Hanvey, J.C., Shimizu, M. & Wells, R.D. (1990) Site-specific inhibition of *Eco*RI restriction/modification enzymes by a DNA triple helix. *Nucleic Acids Res.* **18:** 157–161

Henderson, S.T. & Petes, T.D. (1993) Instability of a plasmid-borne inverted repeat in *Saccharomyces cerevisiae*. *Genetics* **133:** 57–62

Henderson, D. & Weil, J. (1975) Recombination-deficient deletions in bacteriophage λ and their interaction with *chi* mutations. *Genetics* **79:** 143–174

Herman, G.E. & Modrich, P. (1982) *Escherichia coli dam* methylase. *J. Biol. Chem.* **257:** 2605–2612

Herr, W. (1985) Diethyl pyrocarbonate: A chemical probe for secondary structure in negatively supercoiled DNA. *Proc. Natl. Acad. Sci. USA* **82:** 8009–8013

Higgins, C.F., Ames, G.F.-L., Barnes, W.M., Clement, J.M. & Hofnung, M. (1982) A novel intercistronic regulatory element of prokaryotic operons. *Nature (Lond.)* **298:** 760–762

Higgins, C.F., Dorman, C.J. & Ni Bhriain, N. (1990a) Environmental influences on DNA supercoiling: A novel mechanism for the regulation of gene expression. *In:* Drlica, K. & Riley, M. (ed.) The Bacterial Chromosome. pp. 421–432 (American Society for Microbiology, Washington, D.C.)

Higgins, C.F., Dorman, C.J., Stirling, D.A., Waddell, L., Booth, I.R., May, G. & Bremmer, E. (1988) A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. *Cell* **52:** 569–584

Higgins, C.F., Hinton, J.C.D., Hulton, C.S.J., Owen-Hughes, T., Pavitt, G.D. & Seirafi, A. (1990b) Protein H1: A role for chromatin structure in the regulation of bacterial gene expression and virulence? *Mol. Microbiol.* **4:** 2007–2012

Hirota, Y., Yasuda, S., Yamada, M., Nishimura, A., Sugimoto, K., Sugisaki, H., Oka, A. & Takanami, M. (1979) Structural and functional properties of the *Escherichia coli* origin of DNA replication. *Cold Spring Harbor Symp. Quant. Biol.* **43:** 129–138

Hixson, J.E., Wong, T.W. & Clayton, D.A. (1986) Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. *J. Biol. Chem.* **261:** 2384–2390

Hobom, G., Grosschedl, R., Lusky, M., Scherer, G., Schwarz, E. & Kössel, H. (1979) Functional analysis of the replicator structure of lambdoid bacteriophage DNAs. *Cold Spring Harbor Symp. Quant. Biol.* **43:** 165–178

Hoess, R., Wierzbicki, A. & Abremski, K. (1987) Isolation and characterisation of intermediates in site-specific recombination. *Proc. Natl. Acad. Sci. USA* **84:** 6840–6844

Holliday, R. (1964) A mechanism for gene conversion in fungi. *Genet. Res.* **5:** 282–304

Holliday, R. (1989) Untwisting B-Z DNA. *Trends Genet.* **5:** 355–356

Horwitz, M.S.Z. (1989) Transcription regulation *in vitro* by an *E. coli* promoter containing a DNA cruciform in the '–35' region. *Nucleic Acids Res.* **17:** 5537–5545

Horwitz, M.S.Z. & Loeb, L.A. (1988) An *E. coli* promoter that regulates transcription by DNA superhelix-induced cruciform extrusion. *Science (Washington DC)* **241:** 703–705

Howard, F.B., Chen, C., Ross, P.D. & Miles, H.T. (1991) Hairpin formation in the self-complementary dodecamer d-GGTACGCGTACC and derivatives containing GA and IA mispairs. *Biochemistry* **30:** 779–782

Howard-Flanders, P. & Theriot, L. (1966) Mutants of *Escherichia coli* K-12 defective in DNA repair and genetic recombination. *Genetics* **53:** 1137–1150

Hsieh, T. & Wang, J.C. (1975) Thermodynamic properties of superhelical DNAs. Biochemistry **14:** 527–535

Hsu, P.L. & Landy, A. (1984) Resolution of synthetic *att*-site Holliday structures by the integrase protein of bacteriophage λ. *Nature (Lond.)* **311:** 721–726

Hulton, C.S.J., Higgins, C.F. & Sharp, P.M. (1991) ERIC sequences: A novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. *Mol. Microbiol.* **5:** 825–834

Hulton, C.S.J., Seirafi, A., Hinton, J.C.D., Sidebotham, J.M., Waddell, L., Pavitt, G.D., Owen-Hughes, T., Spassky, A., Buc, H. & Higgins, C.F. (1990) Histone-like protein H1 (H-NS), DNA supercoiling and gene expression in bacteria. *Cell* **63:** 631–642

Humayun, Z., Kleid, D. & Ptashne, M. (1977) Sites of contact between λ operators and λ repressor. *Nucleic Acids Res.* **4:** 1595–1607

Hyrien, O. (1989) Large inverted duplications in amplified DNA of mammalian cells form hairpins *in vitro* upon DNA extraction but not *in vivo*. *Nucleic Acids Res.* **17:** 9557–9569

Hyrien, O., Debatisse, M., Buttin, G. & Robert de Saint Vincent, B. (1988) The multicopy appearance of a large inverted duplication and the sequence at the inversion joint suggest a new model for gene amplification. *EMBO (Eur. Mol. Biol. Organ.) J.* **7:** 407–417

Ikeda, H. (1990) DNA topoisomerase-mediated illegitimate recombination. *In:* Cozzarelli, N.R. & Wang, J.C. (ed.) DNA Topology and Its Biological Effects. pp. 341–359, *Cold Spring Harbor Monograph Series* Vol. 20 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York)

Ishiura, M., Hazumi, N., Koide, T., Uchida, T. & Okada, Y. (1989) A *recB recC sbcB recJ* host prevents *recA*-independent deletions in recombinant cosmid DNA propagated in *Escherichia coli*. *J. Bacteriol.* **171:** 1068–1074

Iwasaki, H., Takahagi, M., Shiba, T., Nakata, A. & Shinagawa, H. (1991) *Escherichia coli* RuvC protein is an endonuclease that resolves the Holliday structure. *EMBO (Eur. Mol. Biol. Organ.) J.* **10:** 4381–4389

Jaworski, A., Blaho, J.A., Larson, J.E., Shimizu, M. & Wells, R.D. (1989) Tetracycline promoter mutations decrease non-B DNA structural transitions, negative linking differences and deletions in recombinant plasmids in *Escherichia coli*. *J. Mol. Biol.* **207:** 513–526

Jaworski, A., Hsieh, W., Blaho, J.A., Larson, J.E. & Wells, R.D. (1987) Left-handed DNA *in vivo*. *Science (Washington DC)* **238:** 773–777

Jaworski, A., Zacharias, W., Hsieh, W., Blaho, J.A., Larson, J.E. & Wells, R.D. (1988) *In vivo* existence of left-handed DNA. *Gene (Amst.)* **74:** 215–220

Jayaram, M., Crain, K.L., Parsons, R.L. & Harshey, R.M. (1988) Holliday junctions in FLP recombination: Resolution by step-arrest mutants of FLP protein. *Proc. Natl. Acad. Sci. USA* **85:** 7902–7906

Jensch, F., Kosak, H., Seeman, N.C. & Kemper, B. (1989) Cruciform cutting endonucleases form *Saccharomyces cerevisiae* and phage T4 show conserved reactions with branched DNAs. *EMBO (Eur. Mol. Biol. Organ.) J.* **8:** 4325–4334

Jeyaseelan, R. & Shanmugam, G. (1988) Human placental endonuclease cleaves Holliday junctions. *Biochem. Biophys. Res. Commun.* **156:** 1054–1060

Johnson, A., Meyer, B.J. & Ptashne, M. (1978) Mechanism of action of the *cro* protein of bacteriophage λ. *Proc. Natl. Acad. Sci. USA* **75:** 1783–1787

Johnston, B.H. & Rich, A. (1985) Chemical probes of DNA conformation: Detection of Z-DNA at nucleotide resolution. *Cell* **42:** 713–724

Jones, C.H., Hayward, S.D. & Rawlins, D.R. (1989) Interaction of the lymphocyte-derived Epstein-Barr virus nuclear antigen EBNA-1 with its DNA-binding sites. *J. Virol.* **63:** 101–110

Jones, I.M., Primrose, S.B. & Ehrlich, S.D. (1982) Recombination between short direct repeats in a *Rec*A host. *Mol. & Gen. Genet.* **188:** 486–489

Kaliman, A.V., Kryukov, V.M. & Bayev, A.A. (1988) The nucleotide sequence of the region of bacteriophage T5 early genes D10 – D15. *Nucleic Acids Res.* **16:** 10353–10354

Kallenbach, N.R., Ma, R. & Seeman, N.C. (1983) An immobile nucleic acid junction constructed from oligonucleotides. *Nature (Lond.)* **305:** 829–831

Kallick, D.A. & Wemmer, D.E. (1991) ^{1}H NMR of 5'CGCGTATATACGCG3', a duplex and a four-membered loop. *Nucleic Acids Res.* **19:** 6041–6046

Kang, D.S. & Wells, R.D. (1985) B-Z DNA junctions contain few, if any, nonpaired bases at physiological superhelical densities. *J. Biol. Chem.* **260:** 7783–7790

Kang, C., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. (1992) Crystal structure of four-stranded *Oxytricha* telomeric DNA. *Nature (Lond.)* **356:** 126–131

Karrer, K.M. & Gall, J.G. (1976) The macronuclear ribosomal DNA of *Tetrahymena pyriformis* is a palindrome. *J. Mol. Biol.* **104:** 421–453

Katsura, S., Makishima, F. & Nishimura, H. (1993) Statistical mechanical approach for predicting the transition to non-B structures in supercoiled DNA. *J. Biomol. Struct. & Dyn.* **10:** 639–656

Kazic, T. & Berg, D.E. (1990) Context effects in the formation of deletions in *Escherichia coli*. *Genetics* **126:** 17–24

Kemper, B. & Garabett, M. (1981) Studies on T4-head maturation: 1. Purification and characterisation of gene-*49*-controlled endonuclease. *Eur. J. Biochem.* **115:** 123–131

Kitts, P.A. & Nash, H.A. (1987) Homology-dependent interactions in phage λ site-specific recombination. *Nature (Lond.)* **329:** 346–348

Kleckner, N. (1989) Transposon Tn*10*. *In:* Berg, D.E. & Howe, M.M. (ed.) Mobile DNA. pp. 227–268 (American Society for Microbiology, Washington, D.C.)

Kleff, S., Kemper, B. & Sternglanz, R. (1992) Identification and characterization of yeast mutants and the gene for a cruciform cutting endonuclease. *EMBO (Eur. Mol. Biol. Organ.) J.* **11:** 699–704

Klysik, J. (1992) Cruciform extrusion facilitates intramolecular triplex formation between distal oligopurine·oligopyrimidine tracts: Long range effects. *J. Biol. Chem.* **267:** 17430–17437

Kmiec, E.B. & Holloman, W.K. (1986) Homologous pairing of DNA molecules by Ustilago Rec1 protein is promoted by sequences of Z-DNA. *Cell* **44:** 545–554

Kobayashi, I., Stahl, M.M., Fairfield, F.R. & Stahl, F.W. (1984) Coupling with packaging explains apparent nonreciprocality of Chi-stimulated recombination of bacteriophage lambda by RecA and RecBC functions. *Genetics* **108:** 773–794

Kochanek, S., Renz, D. & Doerfler, W. (1993) Probing DNA-protein interactions *in vitro* with the CpG DNA methyltransferase. *Nucleic Acids Res.* **21:** 2339–2342

Kohwi-Shigematsu, T. & Kohwi, Y. (1992) Detection of non-B-DNA structures at specific sites in supercoiled plasmid DNA and chromatin with haloacetaldehyde and diethyl pyrocarbonate. *In:* Lilley, D.M.J. & Dahlberg, J.E. (ed.) DNA Structures Part B: Chemical and Electrophoretic Analysis of DNA. pp. 155–180, *Methods in Enzymology* Vol. 212 (Academic Press Inc., San Diego)

Kotani, H., Kmiec, E.B. & Holloman, W.K. (1993) Purification and properties of a cruciform DNA binding protein from *Ustilago maydis*. *Chromosoma (Berl.)* **102:** 348–354

Kowalski, D., Natale, D.A. & Eddy, M.J. (1988) Stable DNA unwinding, not breathing, accounts for single-strand-specific nuclease hypersensitivity of specific A+T-rich sequences. *Proc. Natl. Acad. Sci. USA* **85:** 9464–9468

Kraczkiewicz-Dowjat, A. & Fishel, R. (1990) *recB recC*-dependent processing of heteroduplex DNA stimulates recombination of an adjacent gene in *Escherichia coli*. *J. Bacteriol.* **172:** 172–178

Kulkarni, S.K. & Stahl, F.W. (1989) Interaction between the *sbcC* gene of *Escherichia coli* and the *gam* gene of phage λ. *Genetics* **123:** 249–253

Kushner, S.R., Nagaishi, H., Templin, A. & Clark, A.J. (1971) Genetic recombination in *Escherichia coli*: The role of exonuclease I. *Proc. Natl. Acad. Sci. USA* **68:** 824–827

Kutter, E.M. & Wiberg, J.S. (1968) Degradation of cytosine-containing bacterial and bacteriophage DNA after infection of *Escherichia coli* B with bacteriophage T4D wild type and with mutants defective in genes *46*, *47* and *56*. *J. Mol. Biol.* **38:** 395–411

L' Abbé, D., Duhaime, J., Lang, B.F. & Morais, R. (1991) The transcription of DNA in chicken mitochondria initiates from one major bidirectional promoter. *J. Biol. Chem.* **266:** 10844–10850

Lacks, S. & Greenberg, B. (1975) A deoxyribonuclease of *Diplococcus pneumoniae* specific for methylated DNA. *J. Biol. Chem.* **250:** 4060–4066

LaDuca, R.J., Fay, P.J., Chuang, C., McHenry, C.S. & Bambara, R.A. (1983) Site-specific pausing of deoxyribonucleic acid synthesis catalysed by four forms of *Escherichia coli* DNA polymerase III. *Biochemistry* **22:** 5177–5188

Lam, S.T., Stahl, M.M., McMilin, K.D. & Stahl, F.W. (1974) Rec-mediated recombinational hot spot activity in bacteriophage lambda: II. A mutation which causes hot spot activity. *Genetics* **77:** 425–433

Lawn, R.M. (1977) Gene-sized DNA molecules of the *Oxytricha* macronucleus have the same terminal sequence. *Proc. Natl. Acad. Sci. USA* **74:** 4325–4328

Leach, D.R.F. (1987) Site-specific recombination: Making Holliday junctions. *Nature (Lond.)* **329:** 290–291

Leach, D.R.F. & Lindsey, J.C. (1986) *In vivo* loss of supercoiled DNA carrying a palindromic sequence. *Mol. & Gen. Genet.* **204:** 322–327

Leach, D.R.F., Lindsey, J.C. & Okely, E.A. (1987) Genome interactions which influence DNA palindrome mediated instability and inviability in *Escherichia coli*. *J. Cell Sci. Suppl.* **7:** 33–40

Leach, D.R.F., Lloyd, R.G. & Coulson, A.F.W. (1992) The SbcCD Protein of *Escherichia coli* is related to two putative nucleases in the UvrA superfamily of nucleotide-binding proteins. *Genetica (Dordr.)* **87:** 95–100

Leach, D.R.F. & Stahl, F.W. (1983) Viability of λ phages carrying a perfect palindrome in the absence of recombination nucleases. *Nature (Lond.)* **305:** 448–451

Lee, F. & Yanofsky, C. (1977) Transcription termination at the *trp* operon attenuators of *Escherichia coli* and *Salmonella typhimurium*: RNA secondary structure and regulation of termination. *Proc. Natl. Acad. Sci. USA* **74:** 4365–4369

Levinson, G. & Gutman, G.A. (1987) Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4:** 203–221

Lilley, D.M.J. (1980) The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proc. Natl. Acad. Sci. USA* **77:** 6468–6472

Lilley, D.M.J. (1981a) *In vivo* consequences of plasmid topology. *Nature (Lond.)* **292:** 380–382

Lilley, D.M.J. (1981b) Hairpin-loop formation by inverted repeats in supercoiled DNA is a local and transmissible property. *Nucleic Acids Res.* **9:** 1271–1289

Lilley, D.M.J. (1983) Structural perturbation in supercoiled DNA: Hypersensitivity to modification by a single-strand-selective chemical reagent conferred by inverted repeat sequences. *Nucleic Acids Res.* **11:** 3097–3112

Lilley, D.M.J. (1985) The kinetic properties of cruciform extrusion are determined by DNA base-sequence. *Nucleic Acids Res.* **13:** 1443–1465

Lilley, D.M.J. (1986) Bacterial chromatin: A new twist to an old story. *Nature (Lond.)* **320:** 14–15

Lilley, D.M.J. (1988) DNA opens up – supercoiling and heavy breathing. *Trends Genet.* **4:** 111–114

Lilley, D.M.J. (1992a) Probes of DNA structure. *In:* Lilley, D.M.J. & Dahlberg, J.E. (ed.) DNA Structures Part B: Chemical and Electrophoretic Analysis of DNA. pp. 133–139, *Methods in Enzymology* Vol. 212 (Academic Press Inc., San Diego)

Lilley, D.M.J. (1992b) HMG has DNA wrapped up. *Nature (Lond.)* **357:** 282–283

Lilley, D.M.J., Gough, G.W., Hallam, L.R. & Sullivan, K.M. (1985) The physical chemistry of cruciform structures in supercoiled DNA molecules. *Biochimie (Paris)* **67:** 697–706

Lilley, D.M.J. & Hallam, L.R. (1984) Thermodynamics of the ColE1 cruciform: Comparisons between probing and topological experiments using single topoisomers. *J. Mol. Biol.* **180:** 179–200

Lilley, D.M.J. & Kemper, B. (1984) Cruciform-resolvase interactions in supercoiled DNA. *Cell* **36:** 413–422

Lilley, D.M.J. & Markham, A.F. (1983) Dynamics of cruciform extrusion in supercoiled DNA: Use of a synthetic inverted repeat to study conformational populations. *EMBO (Eur. Mol. Biol. Organ.) J.* **2:** 527–533

Lilley, D.M.J. & Palecek, E. (1984) The supercoil-stabilised cruciform of ColE1 is hyper-reactive to osmium tetroxide. *EMBO (Eur. Mol. Biol. Organ.) J.* **3:** 1187–1192

Lin, L. & Meyer, R.J. (1987) DNA synthesis is initiated at two positions within the origin of replication of plasmid R1162. *Nucleic Acids Res.* **15:** 8319–8331

Lindsey, J.C. (1987) Palindrome mediated inviability in *Escherichia coli*. *Ph.D. Thesis*, 159 pp. (University of Edinburgh)

Lindsey, J.C. & Leach, D.R.F. (1989) Slow replication of palindrome-containing DNA. *J. Mol. Biol.* **206:** 779–782

Liu, L.F. & Wang, J.C. (1987) Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* **84:** 7024–7027

Lloyd, R.G. (1991) Linkage distortion following conjugational transfer of *sbcC*$^+$ to *recBC sbcBC* strains of *Escherichia coli*. *J. Bacteriol.* **173:** 5694–5698

Lloyd, R.S. & Augustine, M.L. (1987) High-frequency spontaneous deletion of DNA sequences flanked by direct DNA repeats which also contain an internal palindrome. *Mutat. Res.* **179:** 135–142

Lloyd, R.G. & Buckman, C. (1985) Identification and genetic analysis of *sbcC* mutations in commonly used *recBC sbcB* strains of *Escherichia coli* K-12. *J. Bacteriol.* **164:** 836–844

Lloyd, R.G., Evans, N.P. & Buckman, C. (1987) Formation of recombinant *lacZ*$^+$ DNA in conjugational crosses with a *recB* mutant of *Escherichia coli* K12 depends on *recF*, *recJ*, and *recO*. *Mol. & Gen. Genet.* **209:** 135–141

Lloyd, R.G., Porton, M.C. & Buckman, C. (1988) Effect of *recF*, *recJ*, *recO* and *ruv* mutations on ultraviolet survival and genetic recombination in a *recD* strain of *Escherichia coli* K12. *Mol. & Gen. Genet.* **212:** 317–324

Lloyd, R.G. & Sharples, G.J. (1993a) Dissociation of synthetic Holliday junctions by *E. coli* RecG protein. *EMBO (Eur. Mol. Biol. Organ.) J.* **12:** 17–22

Lloyd, R.G. & Sharples, G.J. (1993b) Processing of recombination intermediates by the RecG and RuvAB proteins of *Escherichia coli*. *Nucleic Acids Res.* **21:** 1719–1725

Lockshon, D. & Galloway, D.A. (1986) Cloning and characterization of ori_{L2}, a large palindromic DNA replication origin of herpes simplex virus type 2. *J. Virol.* **58:** 513–521

Lockshon, D. & Morris, D.R. (1985) Sites of reaction of *Escherichia coli* DNA gyrase on pBR322 *in vivo* as revealed by oxolinic acid-induced plasmid linearization. *J. Mol. Biol.* **181:** 63–74

Lu, M., Guo, Q., Marky, L.A., Seeman, N.C. & Kallenbach, N.R. (1992) Thermodynamics of DNA branching. *J. Mol. Biol.* **223:** 781–789

Lu, M., Guo, Q., Seeman, N.C. & Kallenbach, N.R. (1989) DNase I cleavage of branched DNA molecules. *J. Biol. Chem.* **264:** 20851–20854

Luisi-DeLuca, C., Lovett, S.T. & Kolodner, R.D. (1989) Genetic and physical analysis of plasmid recombination in *recB recC sbcB* and *recB recC sbcA Escherichia coli* K-12 mutants. *Genetics* **122:** 269–278

Lundblad, V. & Kleckner, N. (1985) Mismatch repair mutations of *Escherichia coli* K12 enhance transposon excision. *Genetics* **109:** 3–19

Lundblad, V., Taylor, A.F., Smith, G.R. & Kleckner, N. (1984) Unusual alleles of *recB* and *recC* stimulate excision of inverted repeat transposons Tn*10* and Tn*5*. *Proc. Natl. Acad. Sci. USA* **81:** 824–828

Lyamichev, V.I., Panyutin, I.G. & Frank-Kamenetskii, M.D. (1983) Evidence of cruciform structures in superhelical DNA provided by two-dimensional gel electrophoresis. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* **153:** 298–302

Lyons, S.M. & Schendel, P.F. (1984) Kinetics of methylation in *Escherichia coli* K-12. *J. Bacteriol.* **159:** 421–423

Madsen, C.S., McHugh, K.P. & DeKloet, S.R. (1992) Characterization of a major tandemly repeated DNA sequence (RBMII) prevalent among many species of waterfowl (Anatidae). *Genome* **35:** 1037–1044

Majumdar, R. & Thakur, A.R. (1985) Melting transition of covalently closed DNA with supercoil-induced cruciforms. *Nucleic Acids Res.* **13:** 5883–5893

Mandel, M. & Higa, A. (1970) Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* **53:** 159–162

Marians, K.J. (1992) Prokaryotic DNA replication. *Annu. Rev. Biochem.* **61:** 673–719

Marinus, M.G. (1987) DNA methylation in *Escherichia coli*. *Annu. Rev. Genet.* **21:** 113–131

Martinez, E. & de la Cruz, F. (1990) Genetic elements involved in Tn*21* site-specific integration, a novel mechanism for the dissemination of antibiotic resistance genes. *EMBO (Eur. Mol. Biol. Organ.) J.* **9:** 1275–1281

Maxam, A.M. & Gilbert, W. (1977) A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74:** 560–564

Maxam, A.M. & Gilbert, W. (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. *In:* Grossman, L. & Moldave, K. (ed.) Nucleic Acids. pp. 499–560, *Methods in Enzymology* Vol. 65 (Academic Press Inc., New York)

Mazin, A.V., Kuzimov, A.V., Dianov, G.L. & Salganik, R.I. (1991) Mechanisms of deletion formation in *Escherichia coli* plasmids: II. Deletions mediated by short direct repeats. *Mol. & Gen. Genet.* **228:** 209–214

McClellan, J.A., Boublíková, P., Palecek, E. & Lilley, D.M.J. (1990) Superhelical torsion in cellular DNA responds directly to environmental and genetic factors. *Proc. Natl. Acad. Sci. USA* **87:** 8373–8377

McClellan, J.A. & Lilley, D.M.J. (1987) A two-state conformational equilibrium for alternating $(A\text{-}T)_n$ sequences in negatively supercoiled DNA. *J. Mol. Biol.* **197:** 707–721

McClellan, J.A. & Lilley, D.M.J. (1991) Structural alteration in alternating adenine-thymine sequences in positively supercoiled DNA. *J. Mol. Biol.* **219:** 145–149

McGraw, B.R. & Marinus, M.G. (1980) Isolation and characterization of Dam^+ revertants and suppressor mutations that modify secondary phenotypes of *dam-3* strains of *Escherichia coli* K-12. *Mol. & Gen. Genet.* **178:** 309–315

McMilin, K.D., Stahl, M.M. & Stahl, F.W. (1974) Rec-mediated recombinational hot spot activity in bacteriophage lambda: I. Hot spot activity associated with Spi^- deletions and *bio* substitutions. *Genetics* **77:** 409–423

McMurray, C.T., Wilson, W.D. & Douglass, J.O. (1991) Hairpin formation within the enhancer region of the human enkephalin gene. *Proc. Natl. Acad. Sci. USA* **88:** 666–670

Meijer, M., Beck, E., Hansen, F.G., Bergmans, H.E.N., Messer, W., von Meyenburg, K. & Schaller, H. (1979) Nucleotide sequence of the origin of replication of the *Escherichia coli* K-12 chromosome. *Proc. Natl. Acad. Sci. USA* **76:** 580–584

Menzel, R. & Gellert, M. (1983) Regulation of the genes for *E. coli* DNA gyrase: Homeostatic control of DNA supercoiling. *Cell* **34:** 105–113

Meselson, M. (1972) Formation of hybrid DNA by rotary diffusion during genetic recombination. *J. Mol. Biol.* **71:** 795–798

Meselson, M.S. & Radding, C.M. (1975) A general model for genetic recombination. *Proc. Natl. Acad. Sci. USA* **72:** 358–361

Meyer, R.R. & Laine, P.S. (1990) The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol. Rev.* **54:** 342–380

Meyer-Leon, L., Huang, L., Umlauf, S.W., Cox, M.M. & Inman, R.B. (1988) Holliday intermediates and reaction by-products in FLP protein-promoted site-specific recombination. *Mol. Cell. Biol.* **8:** 3784–3796

Mizuuchi, K., Kemper, B., Hays, J. & Weisberg, R.A. (1982a) T4 endonuclease VII cleaves Holliday structures. *Cell* **29:** 357–365

Mizuuchi, K., Mizuuchi, M. & Gellert, M. (1982b) Cruciform structures in palindromic DNA are favoured by DNA supercoiling. *J. Mol. Biol.* **156:** 229–243

Modrich, P. (1989) Methyl-directed DNA mismatch correction. *J. Biol. Chem.* **264:** 6597–6600

Morales, N.M., Cobourn, S.D. & Müller, U.R. (1990) Effect of *in vitro* transcription on cruciform stability. *Nucleic Acids Res.* **18:** 2777–2782

Morgan, W.D., Bear, D.G. & von Hippel, P.H. (1983a) Rho-dependent termination of transcription: I. Identification and characterization of termination sites for transcription from the bacteriophage λ P_R promoter. *J. Biol. Chem.* **258:** 9553–9564

Morgan, W.D., Bear, D.G. & von Hippel, P.H. (1983b) Rho-dependent termination of transcription: II. Kinetics of mRNA elongation during transcription from the bacteriophage λ P_R promoter. *J. Biol. Chem.* **258:** 9565–9574

Mueller, J.E., Kemper, B., Cunningham, R.P., Kallenbach, N.R. & Seeman, N.C. (1988) T4 endonuclease VII cleaves the crossover strands of Holliday junction analogs. *Proc. Natl. Acad. Sci. USA* **85:** 9441–9445

Müller, U.R. & Fitch, W.M. (1982) Evolutionary selection for perfect hairpin structures in viral DNAs. *Nature (Lond.)* **298:** 582–585

Müller, U.R. & Turnage, M.A. (1986) Insertions of palindromic DNA sequences into the J-F intercistronic region of bacteriophage ϕX174 interfere with normal phage growth. *J. Mol. Biol.* **189:** 285–292

Murchie, A.I.H., Bowater, R., Aboul-ela, F. & Lilley, D.M.J. (1992) Helix opening transitions in supercoiled DNA. *Biochim. Biophys. Acta* **1131:** 1–15

Murchie, A.I.H., Carter, W.A., Portugal, J. & Lilley, D.M.J. (1990) The tertiary structure of the four-way DNA junction affords protection against DNase I cleavage. *Nucleic Acids Res.* **18:** 2599–2606

Murchie, A.I.H., Clegg, R.M., von Kitzing, E., Duckett, D.R., Diekmann, S. & Lilley, D.M.J. (1989) Fluorescence energy transfer shows that the four-way DNA junction is a right-handed cross of antiparallel molecules. *Nature (Lond.)* **341:** 763–766

Murchie, A.I.H. & Lilley, D.M.J. (1987) The mechanism of cruciform formation in supercoiled DNA: Initial opening of central basepairs in salt-dependent extrusion. *Nucleic Acids Res.* **15:** 9641–9654

Murchie, A.I.H. & Lilley, D.M.J. (1989) Base methylation and local DNA helix stability: Effect on the kinetics of cruciform extrusion. *J. Mol. Biol.* **205:** 593–602

Murchie, A.I.H. & Lilley, D.M.J. (1992) Supercoiled DNA and cruciform structures. *In:* Lilley, D.M.J. & Dahlberg, J.E. (ed.) DNA Structures Part A: Synthesis and Physical Analysis of DNA. pp. 158–180, *Methods in Enzymology* Vol. 211 (Academic Press Inc., San Diego)

Murchie, A.I.H., Portugal, J. & Lilley, D.M.J. (1991) Cleavage of a four-way DNA junction by a restriction enzyme spanning the point of strand exchange. *EMBO (Eur. Mol. Biol. Organ.) J.* **10:** 713–718

Nader, W.F., Edlind, T.D., Huettermann, A. & Sauer, H.W. (1985) Cloning of *Physarum* actin sequences in an exonuclease-deficient bacterial host. *Proc. Natl. Acad. Sci. USA* **82:** 2698–2702

Nag, D.K. & Petes, T.D. (1991) Seven-base-pair inverted repeats in DNA form stable hairpins *in vivo* in *Saccharomyces cerevisiae*. *Genetics* **129:** 669–673

Nag, D.K., White, M.A. & Petes, T.D. (1989) Palindromic sequences in heteroduplex DNA inhibit mismatch repair in yeast. *Nature (Lond.)* **340:** 318–320

Nagl, W., Knapp, B. & Bill, O. (1991) The complex satellite DNA of *Tropaeolum majus* L.: Partial characterization of isolated and of cloned restriction fragments. *Ann. Bot. (Lond.)* **67:** 347–355

Naom, I.S., Morton, S.J., Leach, D.R.F. & Lloyd, R.G. (1989) Molecular organization of *sbcC*, a gene that affects genetic recombination and the viability of DNA palindromes in *Escherichia coli* K-12. *Nucleic Acids Res.* **17:** 8033–8045

Naylor, L.H., Lilley, D.M.J. & Van de Sande, J.H. (1986) Stress-induced cruciform formation in a cloned $d(CATG)_{10}$ sequence. *EMBO (Eur. Mol. Biol. Organ.) J.* **5:** 2407–2413

Noirot, P., Bargonetti, J. & Novick, R.P. (1990) Initiation of rolling-circle replication in pT181 plasmid: Initiator protein enhances cruciform extrusion at the origin. *Proc. Natl. Acad. Sci. USA* **87:** 8560–8564

Norrander, J., Kempe, T. & Messing, J. (1983) Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene (Amst.)* **26:** 101–106

Nunes-Düby, S., Matsumoto, L. & Landy, A. (1987) Site-specific recombination intermediates trapped with suicide substrates. *Cell* **50:** 779–788

O'Connor, C.D. & Humphreys, G.O. (1982) Expression of the *Eco*RI restriction-modification system and the construction of positive-selection cloning vectors. *Gene (Amst.)* **20:** 219–229

Ohtsubo, H. & Ohtsubo, E. (1978) Nucleotide sequence of an insertion element, IS1. *Proc. Natl. Acad. Sci. USA* **75:** 615–619

Oka, A., Nomura, N., Morita, M., Sugisaki, H., Kazunori, S. & Takanami, M. (1979) Nucleotide sequence of small ColE1 derivatives: Structure of the regions essential for autonomous replication and colicin E1 immunity. *Mol. & Gen. Genet.* **172:** 151–159

Orlowski, R. & Miller, G. (1991) Single-stranded structures are present within plasmids containing the Epstein-Barr virus latent origin of replication. *J. Virol.* **65:** 677–686

Orr-Weaver, T.L., Szostak, J.W. & Rothstein, R.J. (1981) Yeast transformation: A model system for the study of recombination. *Proc. Natl. Acad. Sci. USA* **78:** 6354–6358

Oxender, D.L., Zurawski, G. & Yanofsky, C. (1979) Attenuation in the *Escherichia coli* tryptophan operon: Role of RNA secondary structure involving the tryptophan codon region. *Proc. Natl. Acad. Sci. USA* **76:** 5524–5528

Palecek, E. (1992a) Probing DNA structure with osmium tetroxide complexes *in vitro*. *In:* Lilley, D.M.J. & Dahlberg, J.E. (ed.) DNA Structures Part B: Chemical and Electrophoretic Analysis of DNA. pp. 139–155, *Methods in Enzymology* Vol. 212 (Academic Press Inc., San Diego)

Palecek, E. (1992b) Probing of DNA structure in cells with osmium tetroxide-2,2'-bipyridine. *In:* Lilley, D.M.J. & Dahlberg, J.E. (ed.) DNA Structures Part B: Chemical and Electrophoretic Analysis of DNA. pp. 305–318, *Methods in Enzymology* Vol. 212 (Academic Press Inc., San Diego)

Panayotatos, N. & Fontaine, A. (1987) A native cruciform DNA structure probed in bacteria by recombinant T7 endonuclease. *J. Biol. Chem.* **262:** 11364–11368

Panayotatos, N. & Wells, R.D. (1981) Cruciform structures in supercoiled DNA. *Nature (Lond.)* **289:** 466–470

Panyutin, I., Klishko, V. & Lyamichev, V. (1984) Kinetics of cruciform formation and stability of cruciform structure in superhelical DNA. *J. Biomol. Struct. & Dyn.* **1:** 1311–1324

Papanicolaou, C. & Ripley, L.S. (1991) An *in vitro* approach to identifying specificity determinants of mutagenesis mediated by DNA misalignments. *J. Mol. Biol.* **221:** 805–821

Park, C., Campbell, J.L. & Goddard, W.A. (1993) Design superiority of palindromic DNA sites for site-specific recognition of proteins: Tests using protein stitchery. *Proc. Natl. Acad. Sci. USA* **90:** 4892–4896

Parniewski, P., Kwinkowski, M., Wilk, A. & Klysik, J. (1990) Dam methyltransferase sites located within the loop region of the oligopurine-oligopyrimidine sequences capable of forming H-DNA are undermethylated *in vivo*. *Nucleic Acids Res.* **18:** 605–611

Parsons, C.A., Tsaneva, I., Lloyd, R.G. & West, S.C. (1992) Interaction of *Escherichia coli* RuvA and RuvB proteins with synthetic Holliday junctions. *Proc. Natl. Acad. Sci. USA* **89:** 5452–5456

Parsons, C.A. & West, S.C. (1988) Resolution of model Holliday junctions by yeast endonuclease is dependent upon homologous DNA sequences. *Cell* **52:** 621–629

Peeters, B.P.H., de Boer, J.H., Bron, S. & Venema, G. (1988) Structural plasmid instability in *Bacillus subtilis*: Effect of direct and inverted repeats. *Mol. & Gen. Genet.* **212:** 450–458

Perlman, S., Phillips, C. & Bishop, J.O. (1976) A study of foldback DNA. *Cell* **8:** 33–42

Perricaudet, M., Fritsch, A., Pettersson, U., Philipson, L. & Tiollais, P. (1977) Excision and recombination of adenovirus DNA fragments in *Escherichia coli*. *Science (Washington DC)* **196:** 208–210

Pettijohn, D.E. & Hodges-Garcia, Y. (1990) Role of HU protein in transitional DNA coiling. *In:* Drlica, K. & Riley, M. (ed.) The Bacterial Chromosome. pp. 241–245 (American Society for Microbiology, Washington, D.C.)

Picksley, S.M., Parsons, C.A., Kemper, B. & West, S.C. (1990) Cleavage specificity of bacteriophage T4 endonuclease VII and bacteriophage T7 endonuclease I on synthetic branch migratable Holliday junctions. *J. Mol. Biol.* **212:** 723–735

Pierce, J.C., Kong, D. & Masker, W. (1991) The effect of the length of direct repeats and the presence of palindromes on deletion between directly repeated DNA sequences in bacteriophage T7. *Nucleic Acids Res.* **19:** 3901–3905

Plasterk, R.H.A. & Van de Putte, P. (1984) Genetic switches by DNA inversions in prokaryotes. *Biochim. Biophys. Acta* **782:** 111–119

Platt, J.R. (1955) Possible separation of intertwined nucleic acid chains by transfer-twist. *Proc. Natl. Acad. Sci. USA* **41:** 181–183

Platt, T. (1986) Transcription termination and the regulation of gene expression. *Annu. Rev. Biochem.* **55:** 339–372

Pognan, F. & Paoletti, C. (1992) Does cruciform DNA provide a recognition signal for DNA-topoisomerase II? *Biochimie (Paris)* **74:** 1019–1023

Pontiggia, A., Negri, A., Beltrame, M. & Bianchi, M.E. (1993) Protein HU binds specifically to kinked DNA. *Mol. Microbiol.* **7:** 343–350

Potter, S.S. (1982) DNA sequence of a foldback transposable element in *Drosophila*. *Nature (Lond.)* **297:** 201–204

Potter, H. & Dressler, D. (1976) On the mechanism of genetic recombination: Electron microscopic observation of recombination intermediates. *Proc. Natl. Acad. Sci. USA* **73:** 3000–3004

Potter, H. & Dressler, D. (1978) *In vitro* system from *Escherichia coli* that catalyses generalized genetic recombination. *Proc. Natl. Acad. Sci. USA* **75:** 3698–3702

Potter, H. & Dressler, D. (1979) DNA recombination: *In vivo* and *in vitro* studies. *Cold Spring Harbor Symp. Quant. Biol.* **43:** 969–985

Pottmeyer, S. & Kemper, B. (1992) T4 endonuclease VII resolves cruciform DNA with nick and counter-nick and its activity is directed by local nucleotide sequence. *J. Mol. Biol.* **223:** 607–615

Prashad, N. & Hosoda, J. (1972) Role of genes *46* and *47* in bacteriophage T4 reproduction. *J. Mol. Biol.* **70:** 617–635

Pulleyblank, D.E., Shure, M., Tang, D., Vinograd, J. & Vosberg, H. (1975) Action of nicking-closing enzyme on supercoiled and nonsupercoiled closed circular DNA: Formation of a Boltzmann distribution of topological isomers. *Proc. Natl. Acad. Sci. USA* **72:** 4280–4284

Raghunathan, G., Jernigan, R.L., Miles, H.T. & Sasisekharan, V. (1991) Conformational feasibility of a hairpin with two purines in the loop. 5'-d-GGTAC*IA*GTACC-3'. *Biochemistry* **30:** 782–788

Rahmouni, A.R. (1992) Z-DNA as a probe for localized supercoiling *in vivo*. *Mol. Microbiol.* **6:** 569–572

Rahmouni, A.R. & Wells, R.D. (1989) Stabilization of Z DNA in vivo by localized supercoiling. *Science (Washington DC)* **246:** 358–363

Rahmouni, A.R. & Wells, R.D. (1992) Direct evidence for the effect of transcription on local DNA supercoiling in vivo. *J. Mol. Biol.* **223:** 131–144

Ream, L.W., Crisona, N.J. & Clark, A.J. (1978) ColE1 plasmid stability in ExoI$^-$ ExoV$^-$ strains of *Escherichia coli* K-12. *In:* Schlessinger, D. (ed.) Microbiology. pp. 78–80 (American Society for Microbiology, Washington, D.C.)

Rentzeperis, D., Alessi, K. & Marky, L.A. (1993) Thermodynamics of DNA hairpins: Contribution of loop size to hairpin stability and ethidium binding. *Nucleic Acids Res.* **21:** 2683–2689

Resnick, M.A. (1976) The repair of double-strand breaks in DNA: A model involving recombination. *J. Theor. Biol.* **59:** 97–106

Rich, A., Nordheim, A. & Wang, A.H.-J. (1984) The chemistry and biology of left-handed Z-DNA. *Annu. Rev. Biochem.* **53:** 791–846

Riley, M. & Anilionis, A. (1978) Evolution of the bacterial genome. *Annu. Rev. Microbiol.* **32:** 519–560

Rinkel, L.J. & Tinoco, I. (1991) A proton NMR study of a DNA dumb-bell structure with hairpin loops of only two nucleotides: d(CACGTGTGTGCGTGCA). *Nucleic Acids Res.* **19:** 3695–3700

Ripley, L.S. (1982) Model for the participation of quasi-palindromic DNA sequences in frameshift mutation. *Proc. Natl. Acad. Sci. USA* **79:** 4128–4132

Ripley, L.S. (1990) Frameshift mutation: Determinants of specificity. *Annu. Rev. Genet.* **24:** 189–213

Ripley, L.S. & Glickman, B.W. (1983) Unique self-complementarity of palindromic sequences provides DNA structural intermediates for mutation. *Cold Spring Harbor Symp. Quant. Biol.* **47:** 851–861

Rosenberg, M., Court, D., Shimatake, H., Brady, C. & Wulff, D.L. (1978) The relationship between function and DNA sequence in an intercistronic regulatory region in phage λ. *Nature (Lond.)* **272:** 414–423

Rosenberg, S.M. & Hastings, P.J. (1991) The split-end model for homologous recombination at double-strand breaks and at Chi. *Biochimie (Paris)* **73:** 385–397

Ruskin, B. & Fink, G.R. (1993) Mutations in *POL1* increase the mitotic instability of tandem inverted repeats in *Saccharomyces cerevisiae*. *Genetics* **133:** 43–56

Sadler, J.R., Betz, J.L., Tecklenburg, M., Goeddel, D.V., Yansura, D.G. & Caruthers, M.H. (1978) Cloning of chemically synthesized lactose operators: II. *Eco*RI-linkered operators. *Gene (Amst.)* **3:** 211–232

Saing, K.M., Orii, H., Tanaka, Y., Yanagisawa, K., Miura, A. & Ikeda, H. (1988) Formation of deletion in *Escherichia coli* between direct repeats located in the long inverted repeats of a cellular slime mold plasmid: Participation of DNA gyrase. *Mol. & Gen. Genet.* **214:** 1–5

Salstrom, J.S. & Szybalski, W. (1978) Coliphage λ *nut*L$^-$: A unique class of mutants defective in the site of gene *N* product utilization for antitermination of leftward transcription. *J. Mol. Biol.* **124:** 195–221

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F. & Petersen, G.B. (1982) Nucleotide sequence of bacteriophage λ DNA. *J. Mol. Biol.* **162:** 729–773

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467

Sawitzke, J.A. & Stahl, F.W. (1992) Phage λ has an analog of *Escherichia coli recO*, *recR* and *recF* genes. *Genetics* **130:** 7–16

Schaaper, R.M. (1989) *Escherichia coli* mutator *mutD5* is defective in the *mutHLS* pathway of DNA mismatch repair. *Genetics* **121**: 205–212

Schaaper, R.M., Danforth, B.N. & Glickman, B.W. (1986) Mechanisms of spontaneous mutagenesis: An analysis of the spectrum of spontaneous mutation in the *Escherichia coli lacI* gene. *J. Mol. Biol.* **189:** 273–284

Schaeffer, F., Yeramian, E. & Lilley, D.M.J. (1989) Long-range structural effects in supercoiled DNA: Statistical thermodynamics reveals a correlation between calculated cooperative melting and contextual influence on cruciform extrusion. *Biopolymers* **28:** 1449–1473

Schlötterer, C. & Tautz, D. (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20:** 211–215

Schmid, C.W., Manning, J.E. & Davidson, N. (1975) Inverted repeat sequences in the *Drosophila* genome. *Cell* **5:** 159–172

Scholten, P.M. & Nordheim, A. (1986) Diethyl pyrocarbonate: A chemical probe for DNA cruciforms. *Nucleic Acids Res.* **14:** 3981–3993

Sen, S., Lahiri, A. & Majumdar, R. (1992) Melting characteristics of highly supercoiled DNA. *Biophys. Chem.* **42:** 229–234

Senior, M.M., Jones, R.A. & Breslauer, K.J. (1988) Influence of loop residues on the relative stabilities of DNA hairpin structures. *Proc. Natl. Acad. Sci. USA* **85:** 6242–6246

Sharples, G.J. & Lloyd, R.G. (1990) A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucleic Acids Res.* **18:** 6503–6508

Sharples, G.J. & Lloyd, R.G. (1993) Location of the *Bacillus subtilis sbcD* gene downstream of *addAB*, the analogues of *E. coli recBC*. *Nucleic Acids Res.* **21:** 2010

Sheflin, L.G. & Kowalski, D. (1985) Altered DNA conformations detected by mung bean nuclease occur in promoter and terminator regions of supercoiled pBR322 DNA. *Nucleic Acids Res.* **13:** 6137–6154

Sherratt, D. (1989) Tn*3* and related transposable elements: Site-specific recombination and transposition. *In:* Berg, D.E. & Howe, M.M. (ed.) Mobile DNA. pp. 163–184 (American Society for Microbiology, Washington, D.C.)

Shiba, T., Iwasaki, H., Nakata, A. & Shinagawa, H. (1991) SOS-inducible DNA repair proteins, RuvA and RuvB, of *Escherichia coli*: Functional interactions between RuvA and RuvB for ATP hydrolysis and renaturation of the cruciform structure in supercoiled DNA. *Proc. Natl. Acad. Sci. USA* **88:** 8445–8449

Shurvinton, C.E., Stahl, M.M. & Stahl, F.W. (1987) Large palindromes in the λ phage genome are preserved in a *rec*$^+$ host by inhibiting λ DNA replication. *Proc. Natl. Acad. Sci. USA* **84:** 1624–1628

Sigal, N. & Alberts, B. (1972) Genetic recombination: The nature of a crossed strand-exchange between two homologous DNA molecules. *J. Mol. Biol.* **71:** 789–793

Silberstein, Z. & Cohen, A. (1987) Synthesis of linear multimers of OriC and pBR322 derivatives in *Escherichia coli* K12: Role of recombination and replication functions. *J. Bacteriol.* **169:** 3131–3137

Sinden, R.R., Broyles, S.S. & Pettijohn, D.E. (1983) Perfect palindromic *lac* operator DNA sequence exists as a stable cruciform structure in supercoiled DNA *in vitro* but not *in vivo*. *Proc. Natl. Acad. Sci. USA* **80:** 1797–1801

Sinden, R.R. & Pettijohn, D.E. (1984) Cruciform transitions in DNA. *J. Biol. Chem.* **259:** 6593–6600

Sinden, R.R. & Ussery, D.W. (1992) Analysis of DNA structure *in vivo* using psoralen photobinding: Measurement of supercoiling, topological domains, and DNA-protein interactions. *In:* Lilley, D.M.J. & Dahlberg, J.E. (ed.) DNA Structures Part B: Chemical and Electrophoretic Analysis of DNA. pp. 319–335, *Methods in Enzymology* Vol. 212 (Academic Press Inc., San Diego)

Sinden, R.R. & Wells, R.D. (1992) DNA structure, mutations and human genetic disease. *Curr. Opin. Biotechnol.* **3:** 612–622

Sinden, R.R., Zheng, G., Brankamp, R.G. & Allen, K.N. (1991) On the deletion of inverted repeated DNA in *Escherichia coli*: Effects of length, thermal stability and cruciform formation *in vivo*. *Genetics* **129:** 991–1005

Singh, J. & Klar, A.J.S. (1992) Active genes in budding yeast display enhanced *in vivo* accessibility to foreign DNA methylases: A novel *in vivo* probe for chromatin structure of yeast. *Genes & Dev.* **6:** 186–196

Singleton, C.K., Klysik, J., Stirdivant, S.M. & Wells, R.D. (1982) Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature (Lond.)* **299:** 312–316

Singleton, C.K. & Wells, R.D. (1982) Relationship between superhelical density and cruciform formation in plasmid pVH51. *J. Biol. Chem.* **257:** 6292–6295

Smith, G.R. (1983) General recombination. *In:* Hendrix, R.W., Roberts, J.W., Stahl, F.W. & Weisberg, R.A. (ed.) Lambda II. pp. 175–209, *Cold Spring Harbor Monograph Series* Vol. 13 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York)

Smith, G.R. (1988a) Homologous recombination in prokaryotes. *Microbiol. Rev.* **52:** 1–28

Smith, G.R. (1988b) Homologous recombination sites and their recognition. *In:* Low, K.B. (ed.) The Recombination of Genetic Material. pp. 115–154 (Academic Press, Inc., San Diego, California)

Smith, G.R., Comb, M., Schultz, D.W., Daniels, D.L. & Blattner, F.R. (1981) Nucleotide sequence of the Chi recombinational hotspot $\chi^{+}D$ in bacteriophage lambda. *J. Virol.* **37:** 336–342

Smith, H.R., Humphreys, G.O., Willshaw, G.A. & Anderson, E.S. (1976) Characterisation of plasmids coding for the restriction-endonuclease *Eco*RI. *Mol. & Gen. Genet.* **143:** 319–325

Smith, G.R. & Stahl, F.W. (1985) Homologous recombination promoted by Chi sites and RecBC enzyme of *Escherichia coli*. *BioEssays* **2:** 244–249

Snyder, M. & Drlica, K. (1979) DNA gyrase on the bacterial chromosome: DNA cleavage induced by oxolinic acid. *J. Mol. Biol.* **131:** 287–302

Sobell, H.M. (1972) Molecular mechanism for genetic recombination. *Proc. Natl. Acad. Sci. USA* **69:** 2483–2487

Stahl, F.W., Stahl, M.M., Malone, R.E. & Crasemann, J.M. (1980) Directionality and nonreciprocality of Chi-stimulated recombination in phage λ. *Genetics* **94:** 235–248

Stary, A. & Sarasin, A. (1992) Molecular analysis of DNA junctions produced by illegitimate recombination in human cells. *Nucleic Acids Res.* **20:** 4269–4274

Staudenbauer, W.L. (1976) Replication of small plasmids in extracts of *Escherichia coli*: Requirement of both DNA polymerases I and III. *Mol. & Gen. Genet.* **149:** 151–158

Stern, M.J., Ames, G.F.-L., Smith, N.H., Robinson, E.C. & Higgins, C.F. (1984) Repetitive extragenic palindromic sequences: A major component of the bacterial genome. *Cell* **37:** 1015–1026

Stern, M.J., Prossnitz, E. & Ames, G.F.-L. (1988) Role of the intercistronic region in post-transcriptional control of gene expression in the histidine transport operon of *Salmonella typhimurium*: Involvement of REP sequences. *Mol. Microbiol.* **2:** 141–152

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. & Inouye, M. (1966) Frameshift mutations and the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **31:** 77–84

Stroynowski, I. & Yanofsky, C. (1982) Transcript secondary structures regulate transcription termination at the attenuator of *S. marcescens* tryptophan operon. *Nature (Lond.)* **298:** 34–38

Sullivan, K.M. & Lilley, D.M.J. (1986) A dominant influence of flanking sequences on a local structural transition in DNA. *Cell* **47:** 817–827

Sullivan, K.M. & Lilley, D.M.J. (1987) Influence of cation size and charge on the extrusion of a salt-dependent cruciform. *J. Mol. Biol.* **193:** 397–404

Sullivan, K.M. & Lilley, D.M.J. (1988) Helix stability and the mechanism of cruciform extrusion in supercoiled molecules. *Nucleic Acids Res.* **16:** 1079–1093

Sullivan, K.M., Murchie, A.I.H. & Lilley, D.M.J. (1988) Long-range structural communication between sequences in supercoiled DNA. *J. Biol. Chem.* **263:** 13074–13082

Summers, D.K. & Sherratt, D.J. (1984) Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability. *Cell* **36:** 1097–1103

Sundquist, W.I. & Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature (Lond.)* **342:** 825–829

Symington, L.S. & Kolodner, R. (1985) Partial purification of an enzyme from *Saccharomyces cerevisiae* that cleaves Holliday junctions. *Proc. Natl. Acad. Sci. USA* **82:** 7247–7251

Syvanen, M., Hopkins, J.D., Griffin, T.J., Liang, T., Ippen-Ihler, K. & Kolodner, R. (1986) Stimulation of precise excision and recombination by conjugal proficient F' plasmids. *Mol. & Gen. Genet.* **203:** 1–7

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. & Stahl, F.W. (1983) The double-strand-break repair model for recombination. *Cell* **33:** 25–35

Taylor, A.F. (1988) RecBCD enzyme of *Escherichia coli*. *In:* Kucherlapati, R. & Smith, G.R. (ed.) Genetic Recombination. pp. 231–263 (American Society for Microbiology, Washington, DC)

Taylor, A.F. (1992) Movement and resolution of Holliday junctions by enzymes from *E. coli*. *Cell* **69:** 1063–1065

Taylor, A.F. & Smith, G.R. (1990) Action of RecBCD enzyme on cruciform DNA. *J. Mol. Biol.* **211:** 117–134

Tenen, D.G., Livingston, D.M., Wang, S. & Martin, R.G. (1983) Effect of a stem-loop structure within the SV40 replication origin upon SV40 T antigen binding to origin region sequences. *Cell* **34:** 629–639

Thaler, D.S. & Stahl, F.W. (1988) DNA double-chain breaks in recombination of phage λ and of yeast. *Annu. Rev. Genet.* **22:** 169–197

Thaler, D.S., Stahl, M.M. & Stahl, F.W. (1987) Tests of the double-strand-break repair model for Red-mediated recombination of phage λ and plasmid λdv. *Genetics* **116:** 501–511

Thiyagarajan, M.M., Kotani, H., Holloman, W.K. & Kmiec, E.B. (1993) DNA relaxation mediated by *Ustilago maydis* type I topoisomerase; modulation by chromatin associated proteins. *Biochim. Biophys. Acta* **1173:** 155–164

Thompson, B.J., Escarmis, C., Parker, B., Slater, W.C., Doniger, J., Tessman, I. & Warner, R.C. (1975) Figure-8 configuration of dimers of S13 and ϕX174 replicative form DNA. *J. Mol. Biol.* **91:** 409–419

Tijan, R. (1978) Protein-DNA interactions at the origin of simian virus 40 DNA replication. *Cold Spring Harbor Symp. Quant. Biol.* **43:** 655–662

Trinh, T.Q. & Sinden, R.R. (1991) Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. *Nature (Lond.)* **352:** 544–547

Trinh, T.Q. & Sinden, R.R. (1993) The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in *Escherichia coli*. *Genetics* **134:** 409–422

Truett, M.A., Jones, R.S. & Potter, S.S. (1981) Unusual structure of the FB family of transposable elements in *Drosophila*. *Cell* **24:** 753–763

Tsaneva, I.R., Müller, B. & West, S.C. (1992) ATP-dependent branch migration of Holliday junctions promoted by the RuvA and RuvB proteins of *E. coli*. *Cell* **69:** 1171–1180

Tsaneva, I.R., Müller, B. & West, S.C. (1993) RuvA and RuvB proteins of *Escherichia coli* exhibit DNA helicase activity *in vitro*. *Proc. Natl. Acad. Sci. USA* **90:** 1315–1319

Unger, R.C. & Clark, A.J. (1972) Interaction of the recombination pathways of bacteriophage λ and its host *Escherichia coli* K12: Effects on exonuclease V activity. *J. Mol. Biol.* **70:** 539–548

Unger, R.C., Echols, H. & Clark, A.J. (1972) Interaction of the recombination pathways of bacteriophage λ and host *Escherichia coli*: Effects on λ recombination. *J. Mol. Biol.* **70:** 531–537

Ussery, D.W., Hoepfner, R.W. & Sinden, R.R. (1992) Probing DNA structure with psoralen *in vitro*. *In:* Lilley, D.M.J. & Dahlberg, J.E. (ed.) DNA Structures Part B: Chemical and Electrophoretic Analysis of DNA. pp. 242–262, *Methods in Enzymology* Vol. 212 (Academic Press Inc., San Diego)

Van de Ven, F.J.M. & Hilbers, C.W. (1988) Nucleic acids and nuclear magnetic resonance. *Eur. J. Biochem.* **178:** 1–38

Vardimon, L. & Rich, A. (1984) In Z-DNA the sequence G-C-G-C is neither methylated by *Hha*I methyltransferase nor cleaved by *Hha*I restriction endonuclease. *Proc. Natl. Acad. Sci. USA* **81:** 3268–3272

Varga-Weisz, P., van Holde, K. & Zlatanova, J. (1993) Preferential binding of histone H1 to four-way helical junction DNA. *J. Biol. Chem.* **268:** 20699–20700

Veaute, X. & Fuchs, R.P.P. (1993) Greater susceptibility to mutations in lagging strand of DNA replication in *Escherichia coli* than in leading strand. *Science (Washington DC)* **261:** 598–600

Vinograd, J., Lebowitz, J., Radloff, R., Watson, R. & Laipis, P. (1965) The twisted circular form of polyoma viral DNA. *Proc. Natl. Acad. Sci. USA* **53:** 1104–1111

Vogelstein, B. & Gillespie, D. (1979) Preparative and analytical purification of DNA from agarose. *Proc. Natl. Acad. Sci. USA* **76:** 615–619

Vologodskii, A.V. & Frank-Kamenetskii, M.D. (1982) Theoretical study of cruciform states in superhelical DNAs. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* **143:** 257–260

Vologodskii, A.V. & Frank-Kamenetskii, M.D. (1983) The relaxation time for a cruciform structure in superhelical DNA. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* **160:** 173–176

Vologodskii, A.V., Lukashin, A.V., Anshelevich, V.V. & Frank-Kamenetskii, M.D. (1979) Fluctuations in superhelical DNA. *Nucleic Acids Res.* **6:** 967–982

von Kitzing, E., Lilley, D.M.J. & Diekmann, S. (1990) The stereochemistry of a four-way DNA junction: A theoretical study. *Nucleic Acids Res.* **18:** 2671–2683

Vovis, G.F. & Lacks, S. (1977) Complementary action of restriction enzymes Endo R·*Dpn*I and Endo R·*Dpn*II on bacteriophage f1 DNA. *J. Mol. Biol.* **115:** 525–538

Waga, S., Mizuno, S. & Yoshida, M. (1990) Chromosomal protein HMG1 removes the transcriptional block caused by the cruciform in supercoiled DNA. *J. Biol. Chem.* **265:** 19424–19428

Wagner, R.E. & Radman, M. (1975) A mechanism for initiation of genetic recombination. *Proc. Natl. Acad. Sci. USA* **72:** 3619–3622

Waldman, A.S. & Liskay, R.M. (1988) Resolution of synthetic Holliday structures by an extract of human cells. *Nucleic Acids Res.* **16:** 10249–10266

Walker, P.M.B. & McLaren, A. (1965) Fractionation of mouse deoxyribonucleic acid on hydroxyapatite. *Nature (Lond.)* **208:** 1175–1179

Walz, A. & Pirrotta, V. (1975) Sequence of the P_R promoter of phage λ. *Nature (Lond.)* **254:** 118–121

Wang, J.C. (1979) Helical repeat of DNA in solution. *Proc. Natl. Acad. Sci. USA* **76:** 200–203

Wang, M.X. & Church, G.M. (1992) A whole genome approach to *in vivo* DNA-protein interactions in *E. coli*. *Nature (Lond.)* **360:** 606–610

Wang, J.C., Peck, L.J. & Becherer, K. (1983) DNA supercoiling and its effects on DNA structure and function. *Cold Spring Harbor Symp. Quant. Biol.* **47:** 85–91

Wang, P., Projan, S.J., Henriquez, V. & Novick, R.P. (1993) Origin recognition specificity in pT181 plasmids is determined by a functionally asymmetric palindrome DNA element. *EMBO (Eur. Mol. Biol. Organ.) J.* **12:** 45–52

Ward, P. & Berns, K.I. (1991) *In vitro* rescue of an integrated hybrid adeno-associated virus/simian virus 40 genome. *J. Mol. Biol.* **218:** 791–804

Ward, G.K., McKenzie, R., Zannis-Hadjopoulos, M. & Price, G.B. (1990) The dynamic distribution and quantification of DNA cruciforms in eukaryotic nuclei. *Exp. Cell Res.* **188:** 235–246

Ward, G.K., Shihab-el-Deen, A., Zannis-Hadjopoulos, M. & Price, G.B. (1991) DNA cruciforms and the nuclear supporting structure. *Exp. Cell Res.* **195:** 92–98

Warren, G.J. & Green, R.L. (1985) Comparison of physical and genetic properties of palindromic DNA sequences. *J. Bacteriol.* **161:** 1103–1111

Watson, J.D. & Crick, F.H.C. (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature (Lond.)* **171:** 737–738

Weaver, D.T. & DePamphilis, M.L. (1984) The role of palindromic and non-palindromic sequences in arresting DNA synthesis *in vitro* and *in vivo*. *J. Mol. Biol.* **180:** 961–986

Weir, H.M., Kraulis, P.J., Hill, C.S., Raine, A.R.C., Laue, E.D. & Thomas, J.O. (1993) Structure of the HMG box motif in the B-domain of HMG1. *EMBO (Eur. Mol. Biol. Organ.) J.* **12:** 1311–1319

Weller, S.K., Spadaro, A., Schaffer, J.E., Murray, A.W., Maxam, A.M. & Schaffer, P.A. (1985) Cloning, sequencing and functional analysis of ori_L, a herpes simplex virus type 1 origin of DNA synthesis. *Mol. Cell. Biol.* **5:** 930–942

Wells, R.D. (1988) Unusual DNA structures. *J. Biol. Chem.* **263:** 1095–1098

Wemmer, D.E., Wand, A.J., Seeman, N.C. & Kallenbach, N.R. (1985) NMR analysis of DNA junctions: Imino proton NMR studies of individual arms and intact junction. *Biochemistry* **24:** 5745–5749

Wertman, K.F., Wyman, A.R. & Botstein, D. (1986) Host/vector interactions which affect the viability of recombinant phage lambda clones. *Gene (Amst.)* **49:** 253–262

Wesley, R.D. (1975) Inverted repetitious sequences in the macronuclear DNA of hypotrichous ciliates. *Proc. Natl. Acad. Sci. USA* **72:** 678–682

West, S.C. (1992) Enzymes and molecular mechanisms of genetic recombination. *Annu. Rev. Biochem.* **61:** 603–640

West, S.C. & Connolly, B. (1992) Biological roles of the *Escherichia coli* RuvA, RuvB and RuvC proteins revealed. *Mol. Microbiol.* **6:** 2755–2759

West, S.C., Countryman, J.K. & Howard-Flanders, P. (1983) Enzymatic formation of biparental figure-eight molecules from plasmid DNA and their resolution in *E. coli*. *Cell* **32:** 817–829

West, S.C. & Körner, A. (1985) Cleavage of cruciform DNA structures by an activity from *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **82:** 6445–6449

Weston-Hafer, K. & Berg, D.E. (1989) Palindromy and the location of deletion endpoints in *Escherichia coli*. *Genetics* **121:** 651–658

Weston-Hafer, K. & Berg, D.E. (1991a) Limits to the role of palindromy in deletion formation. *J. Bacteriol.* **173:** 315–318

Weston-Hafer, K. & Berg, D.E. (1991b) Deletions in plasmid pBR322: Replication slippage involving leading and lagging strands. *Genetics* **127:** 649–655

Whitehouse, H.L.K. (1963) A theory of crossing-over by means of hybrid deoxyribonucleic acid. *Nature (Lond.)* **199:** 1034–1040

Williams, W.L. & Müller, U.R. (1987) Effects of palindrome size and sequence on genetic stability in the bacteriophage ϕX174 genome. *J. Mol. Biol.* **196:** 743–755

Wilson, D.A. & Thomas, C.A. (1974) Palindromes in chromosomes. *J. Mol. Biol.* **84:** 115–144

Wohlrab, F. (1992) Enzyme probes *in vitro*. *In:* Lilley, D.M.J. & Dahlberg, J.E. (ed.) DNA Structures Part B: Chemical and Electrophoretic Analysis of DNA. pp. 294–301, *Methods in Enzymology* Vol. 212 (Academic Press Inc., San Diego)

Wood, W.B. (1966) Host specificity of DNA produced by *Escherichia coli*: Bacterial mutations affecting the restriction and modification of DNA. *J. Mol. Biol.* **16:** 118–133

Wu, H. & Crothers, D.M. (1984) The locus of sequence-directed and protein-induced DNA bending. *Nature (Lond.)* **308:** 509–513

Wu, H., Shyy, S., Wang, J.C. & Liu, L.F. (1988) Transcription generates positively and negatively supercoiled domains in the template. *Cell* **53:** 433–440

Wyman, A.R., Wertman, K.F., Barker, D., Helms, C. & Petri, W.H. (1986) Factors which equalize the representation of genome segments in recombinant libraries. *Gene (Amst.)* **49:** 263–271

Wyman, A.R., Wolfe, L.B. & Botstein, D. (1985) Propagation of some human DNA sequences in bacteriophage λ vectors requires mutant *Escherichia coli* hosts. *Proc. Natl. Acad. Sci. USA* **82:** 2880–2884

Xodo, L.E., Manzini, G., Quadrifoglio, F., Van der Marel, G.A. & van Boom, J.H. (1988) Oligodeoxynucleotide folding in solution: Loop size and stability of B-hairpins. *Biochemistry* **27:** 6321–6326

Xodo, L.E., Manzini, G., Quadrifoglio, F., Van der Marel, G.A. & van Boom, J. (1991) DNA hairpin loops in solution. Correlation between primary structure, thermostability and reactivity with single-strand-specific nuclease from mung bean. *Nucleic Acids Res.* **19:** 1505–1511

Yang, Y. & Ames, G.F.-L. (1988) DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. *Proc. Natl. Acad. Sci. USA* **85:** 8850–8854

Yang, Y. & Ames, G.F.-L. (1990) The family of repetitive extragenic palindromic sequences: Interaction with DNA gyrase and histonelike protein HU. *In:* Drlica, K. & Riley, M. (ed.) The Bacterial Chromosome. pp. 211–225 (American Society for Microbiology, Washington, D.C.)

Yang, R.C.A. & Wu, R. (1979) BK virus DNA: Complete nucleotide sequence of a human tumor virus. *Science (Washington DC)* **206:** 456–462

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene (Amst.)* **33:** 103–119

Yoshimura, H., Yoshino, T., Hirose, T., Nakamura, Y., Higashi, M., Hase, T., Yamaguchi, K., Hirokawa, H. & Masamune, Y. (1986) Biological characteristics of palindromic DNA (II). *J. Gen. Appl. Microbiol.* **32:** 393–404

Zacharias, W. (1992) DNA methylation *in vivo*. *In:* Lilley, D.M.J. & Dahlberg, J.E. (ed.) DNA Structures Part B: Chemical and Electrophoretic Analysis of DNA. pp. 336–346, *Methods in Enzymology* Vol. 212 (Academic Press Inc., San Diego)

Zacharias, W., Larson, J.E., Kilpatrick, M.W. & Wells, R.D. (1984) *Hha*I methylase and restriction endonuclease as probes for B to Z DNA conformational changes in d(GCGC) sequences. *Nucleic Acids Res.* **12:** 7677–7692

Zannis-Hadjopoulos, M., Frappier, L., Khoury, M. & Price, G.B. (1988) Effect of anti-cruciform DNA monoclonal antibodies on DNA replication. *EMBO (Eur. Mol. Biol. Organ.) J.* **7:** 1837–1844

Zheng, G., Kochel, T., Hoepfner, R.W., Timmons, S.E. & Sinden, R.R. (1991) Torsionally tuned cruciform and Z-DNA probes for measuring unrestrained supercoiling at specific sites in DNA of living cells. *J. Mol. Biol.* **221:** 107–129

Zheng, G. & Sinden, R.R. (1988) Effects of base composition at the centre of inverted repeated DNA sequences on cruciform transitions in DNA. *J. Biol. Chem.* **263:** 5356–5361

Zuker, M. & Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9:** 133–148

Zyskind, J.W., Cleary, J.M., Brusilow, W.S.A., Harding, N.E. & Smith, D.W. (1983) Chromosomal replication origin from the marine bacterium *Vibrio harveyi* functions in *Escherichia coli*: *oriC* consensus sequence. *Proc. Natl. Acad. Sci. USA* **80:** 1164–1168

APPENDIX

DNA Sequence of 462 bp Palindrome in λ DRL133 and DRL167

Nucleotide sequence of the 462 bp palindrome in λ DRL133 and DRL167. The construction of the palindrome is described in the text, but is essentially an inverted duplication of the DNA fragment between the *Sac*I site at 25881 bp and the *Eco*RI site at 26104 bp of the wild-type λ sequence (Sanger *et al.*, 1982). It is a derivative of the 571 bp palindrome in λ DRL116, which has been verified by DNA sequencing (J. Lindsey, Ph.D. Thesis, 1987). The restriction endonuclease target sites described in the text are shown in bold.

*Eco*RI
001 A**GAATTC**ATTTCAGCATTTATTGGTTGTATGAGAGTAGATAGAAAAAGAC 050

051 AACTCTGGCTTGAAGCTATCAAAAAACTAAGTAGTGATGAAAACTTTTCA 100

101 AATATGGAACTCATCAGCCTCATTTCTAAATATGAAGAGTTAAGACGTAA 150

151 TGAACCACAGATTCAAGTGGACGATGATAAATTCACTAAATTGTTTTATG 200

*Taq*I *Sac*I *Taq*I
201 ACAATATCCAGAAATATCTGCT**TCGA**AT**GAGCTC**AT**TCGA**AGCAGATATT 250

251 TCTGGATATTGTCATAAAACAATTTAGTGAATTTATCATCGTCCACTTGA 300

301 ATCTGTGGTTCATTACGTCTTAACTCTTCATATTTAGAAATGAGGCTGAT 350

351 GAGTTCCATATTTGAAAAGTTTTCATCACTACTTAGTTTTTTGATAGCTT 400

401 CAAGCCAGAGTTGTCTTTTTCTATCTACTCTCATACAACCAATAAATGCT 450

*Eco*RI
451 GAAAT**GAATTC**T 462